

ESTIMATION OF MULTIPLE POISSON MEANS UNDER ENTROPY LOSS

BY

MING-CHUNG YANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1986

To my brother

ACKNOWLEDGMENTS

I would like very much to thank Dr. Malay Ghosh for being my dissertation advisor and proposing the topics. Specifically, I would like to thank him for his enormous patience, encouragement and guidance. Besides, I would like to thank Dr. Murali Rao, Dr. Ronald Randles, Dr. Andrew Rosalsky, and Dr. Jeffrey Witmer for serving on my committee. Also, I would like to thank Dr. Alan Agresti for his encouragement and help during my study. Finally, I would like to thank the Department of Statistics for providing a very congenial research atmosphere.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	iii
ABSTRACT.....	vi
CHAPTERS	
ONE	INTRODUCTION.....1
	1.1 Background.....1
	1.2 Estimation of Several Poisson Means....3
	1.3 The Subject of This Research.....5
TWO	ADMISSIBLE, BAYES, HIERARCHICAL BAYES AND RIDGE ESTIMATORS.....11
	2.1 Introduction.....11
	2.2 The Admissibility Results.....16
	2.3 Hierarchical Bayes Estimators.....26
	2.4 Monte Carlo Simulations.....28
	2.5 Ridge Estimators of Several Independent Poisson Means.....32
	2.6 Proof of Lemma 2.2.5.....37
THREE	COMPROMISE BETWEEN GENERALIZED BAYES AND BAYES ESTIMATORS.....42
	3.1 Introduction.....42
	3.2 The Risk Performance of $\delta_{M,\alpha,k}^C$ 47
	3.3 The RSL of $\delta_{M,\alpha,k}^C$ 57
FOUR	SUMMARY.....64

BIBLIOGRAPHY.....	65
BIOGRAPHICAL SKETCH.....	69

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ESTIMATION OF MULTIPLE POISSON MEANS UNDER ENTROPY LOSS

BY

MING-CHUNG YANG

August 1986

Chairman: Malay Ghosh
Major Department: Statistics

This dissertation is concerned with simultaneous estimation of Poisson means under entropy loss instead of the usual quadratic loss.

The characterization of admissible linear estimators of multiple Poisson parameters under entropy loss will be presented. Estimators dominating some of the available estimators are given. Hierarchical Bayes estimators are generated, and conditions under which they dominate the available estimators are also given. Monte Carlo simulations are undertaken to indicate the extent of the risk dominance. Further, some properties of ridge estimators of multiple Poisson means under entropy loss are studied.

Compromise estimators between the generalized Bayes and Bayes estimators with respect to conjugate gamma priors under entropy loss are proposed. The proposed compromise estimators are compared with some suitable generalized Bayes and Bayes estimators in terms of their frequentist risk performance. Also the proposed compromise estimators are compared with some admissible generalized Bayes estimators undertaken on the basis of their Bayes risk performance.

CHAPTER ONE INTRODUCTION

1.1 Background

In his seminal 1955 paper, Stein discovered the surprising phenomenon that for estimating p (≥ 3) independent normal means $(\theta_1, \dots, \theta_p) = \underline{\theta}$ simultaneously, there is a better estimator of $\underline{\theta}$ than the sample mean under mean squared error loss. The sample mean in this problem is the maximum likelihood (ML) and uniformly minimum variance unbiased estimator (UMVUE) of $\underline{\theta}$. Later, an explicit estimator dominating the sample means was proposed by James and Stein (1961). This is the so-called James-Stein estimator which shrinks the sample mean towards zero. In practice, however, the positive-part James-Stein estimator is used as it dominates the James-Stein estimator and prevents overshrinking. However, as shown by Brown (1971), both of these estimators are inadmissible. Efron and Morris (1973) argued for the use of positive-part James-Stein estimator since this estimator can not be substantially improved on. Proper Bayes admissible estimators dominating the sample mean were given by Strawderman (1971). more

generally, several improved estimators of a multivariate normal mean have been derived under the specific covariance matrix which may be known or unknown. In this situation, the loss function used is usually quadratic.

On the other hand, there has been considerable interest in simultaneous estimation of non-location parameters as well as means and other parameters of interest for the nonnormal problem. Johnson (1971) considered simultaneous estimation of binomial and multinomial probabilities under squared error loss and showed that there was no Stein-effect in these cases. He pointed out though that admissibility of the sample proportions in such situations was primarily due to the fact that they could not be improved when the θ_1 's were zeroes or ones. This phenomenon has been referred to as the tyranny of the boundary of the parameter space.

So far we have discussed only the squared error or quadratic loss. The entropy loss, i.e. when the loss is defined as the entropy or the Kullback-Leibler distance between two populations, was first introduced in James and Stein (1961) for estimation of the multinormal variance-covariance matrix. Later, the same loss was considered in Brown (1968), Haff (1977, 1979, 1980, 1982), and Dey and Srinivasan (1985) for estimating either the multinormal variance-covariance matrix or its inverse. Dey, Ghosh and

Srinivasan (1985) considered the entropy loss for simultaneous estimation of p independent gamma scale parameters or their reciprocals, while Ighodaro and Santner (1982) and Ighodaro, Santner and Brown (1982) considered entropy loss for simultaneous estimation of independent binomial and multinomial proportions.

The emphasis of this thesis is simultaneous estimation of Poisson means. We review briefly the related literature in the next section.

1.2 Estimation of Several Poisson Means

Assume $X_1, \dots, X_p$ are independent and X_i is distributed Poisson with mean θ_i where $\theta_i \in (0, \infty)$ is unknown for each $i = 1, \dots, p$. In estimating $\underline{\theta} = (\theta_1, \dots, \theta_p)$ by $\underline{a} = (a_1, \dots, a_p)$, consider the loss function

$$L_{\underline{m}}(\underline{\theta}, \underline{a}) = \sum_{i=1}^p \theta_i^{-m_i} (\theta_i - a_i)^2$$

where the m_i 's are known nonnegative integers which reflect the relative severity of misestimation of the different components. Let $\underline{m} = (m_1, \dots, m_p)$, and $\underline{X} = (X_1, \dots, X_p)$. The usual estimator of $\underline{\theta}$ is $\underline{X}$ which is the MLE and the UMVUE. This estimator is also a minimax estimator. Without loss of generality, we can restrict ourselves to one observation

from each population, since the minimal sufficient statistic in this case is the vector of sample sums, and its components are also independent Poissons.

Peng (1975) showed that $\bar{X}$ is inadmissible under the squared error loss L_0 when $p \geq 3$. Given the normalized squared error loss L_1 , Clevenson and Zidek (1975) first proved the inadmissibility of the usual estimator $\bar{X}$ of θ for $p > 2$. Indeed, they proposed a class of estimators of θ of the form

$$\hat{\delta}(\bar{X}) = (1 - (\beta + p - 1) / (\sum_{i=1}^p X_i + \beta + p - 1)) \bar{X},$$

$0 \leq \beta \leq p-1$ and $p \geq 2$, which has uniformly smaller risk than that of $\bar{X}$ when the loss is L_1 . Such a $\hat{\delta}(\bar{X})$ is also a minimax estimator of θ under L_1 , since $\bar{X}$ itself is minimax. These estimators are called Clevenson-Zidek-type estimators. Tsui (1986) has shown that the superiority of the CZ-type estimators over the usual estimator $\bar{X}$ given the loss L_1 is robust with respect to the underlying distributions belonging to a large class of discrete distributions including but not limited to the Poisson distribution. Ghosh and Parsian (1981) and Ghosh (1983) have obtained some Bayes (with respect to certain two stage priors) and empirical Bayes estimators dominating $\bar{X}$, preserving thereby the minimaxity property. Ghosh (1983)

proposed a hierarchical Bayes analysis by using two stage priors. The results of Ghosh (1983) generalized the earlier results of Ghosh and Parsian (1981) to a wider class. Both these papers generate a class of proper Bayes admissible estimators. Also, Ghosh (1983) came up with an interesting empirical Bayes interpretation of the Hierarchical Bayes estimators.

The above estimators, while improving uniformly on $\bar{X}$, shrink the usual estimator $\bar{X}$ towards the origin. However, there are situations when $\bar{X}$ is thought to be away from the origin, and estimators shrinking only towards the origin will not offer much improvement. In such cases, one should shrink towards points other than zero. Hudson and Tsui (1981) considered shrinking towards a data based point under squared error loss. Ghosh, Hwang and Tsui (1983) considered the simultaneous estimation for the general discrete exponential family. They came up with some estimators shrinking the usual estimator $\bar{X}$ towards an arbitrarily prechosen point or a data based point, rather than zero.

The simultaneous estimation of several Poisson means can be applied to many areas. Estimators improving on $\bar{X}$ have been applied to oil-well discovery by Clevenston and Zidek (1975), crime rate estimation by Rolph, Chaiken and Houchens (1981), quality assurance problems by Hoadley

(1981), and error rate estimation in audit sampling by Matsumura and Tsui (1982).

1.3 The Subject of This Research

In this dissertation, we consider simultaneous estimation of multiple Poisson means under entropy loss instead of the squared error loss or relative squared error loss. To motivate this, first consider the model where observations $X_1, \dots, X_p$ are independent and normally distributed with respect to means $\theta_1, \dots, \theta_p$ and common known variance σ^2 under the quadratic loss given by

$$\begin{aligned} L_1(\underline{\theta}, \underline{a}) &= (\underline{\theta} - \underline{a})' \underline{\Sigma}^{-1} (\underline{\theta} - \underline{a}) \\ &= \sum_{i=1}^p (\theta_i - a_i)^2 / \sigma^2 \end{aligned}$$

where $\underline{\Sigma} = \sigma^2 \mathbf{I}_p$. On the other hand, we consider the loss as the entropy distance (Kullback-Leiber information number) between two distributions, each of which is the joint distribution of p independent normal variables. Kullback (1959) described this quantity as the mean information from the likelihood function $f(\underline{X}; \underline{\theta})$ against $f(\underline{X}; \underline{a})$. The resulting entropy loss is defined as

$$L_2(\underline{\theta}, \underline{a}) = E_{\underline{\theta}}(\log(f(\underline{X}; \underline{\theta})/f(\underline{X}; \underline{a})))$$

$$\begin{aligned}
&= E_{\theta} \left\{ \sum_{i=1}^p ((X_i - a_i)^2 - (X_i - \theta_i)^2) / 2\sigma^2 \right\} \\
&= \sum_{i=1}^p (\theta_i - a_i)^2 / 2\sigma^2 \\
&= L_1(\underline{\theta}, \underline{a}) / 2.
\end{aligned}$$

Consequently, the above entropy loss is equivalent to the quadratic loss when the underlying distribution is normal. For the estimation of multiple Poisson means, the entropy loss is quite different from the quadratic loss. That is one motivation for this research. The entropy loss in the Poisson case is given by

$$\begin{aligned}
L(\underline{\theta}, \underline{a}) &= E_{\theta} \left(\log \left\{ \left(\prod_{i=1}^p \exp(-\theta_i) \theta_i^{X_i} / X_i! \right) / \left(\prod_{i=1}^p \exp(-a_i) a_i^{X_i} / X_i! \right) \right\} \right) \\
&= \sum_{i=1}^p (a_i - \theta_i - \theta_i \log \frac{a_i}{\theta_i}) \\
&= \sum_{i=1}^p \theta_i \left(\frac{a_i}{\theta_i} - \log \frac{a_i}{\theta_i} - 1 \right). \tag{1.3.1}
\end{aligned}$$

In the above, we adopt the convention of interpreting $0/0$ as 1 and $0 \log 0$ as 0.

In Chapter Two, we characterize admissible linear estimators of Poisson means with the form $\underline{X}\underline{C} + \underline{b}$, where $\underline{C}$ is a known diagonal matrix and $\underline{b}$ is a known vector, under entropy loss. Also, we present some interesting admissible proper Bayes estimators by using hierarchical priors. We point out also in this chapter that Bayes estimators can be viewed as ridge estimators. Chapter Three considers

compromise estimators between generalized Bayes and Bayes estimators.

Most of the inadmissibility results for the one-parameter discrete exponential family were obtained by first deriving some difference inequalities using Stein' (1973) idea, and then solving these inequalities by guess work. The solutions of these inequalities were used to construct the improved estimators which usually shift the usual estimators towards the zero. For our case, a new difference inequality will be presented and solved in detail in Section 2 of Chapter Two. For the proof of admissibility, our proof employs Blyth's (1951) technique with a sequence of non-conjugate priors.

Hoerl and Kennard (1970) originally introduced ridge estimators of β for multiple linear regression model

$$Y = X\beta + \varepsilon.$$

The ridge estimators are given by

$$\hat{\beta}_k = (X'X + kI)^{-1}X'Y,$$

where $k (\geq 0)$ is a constant. Assuming normality of error and full rank of the matrix of independent variables X , these ridge estimators can be shown to dominate the least square estimator of β given by

$$\hat{\beta}_0 = (X'X)^{-1}X'Y$$

under squared error loss. Furthermore, the class of ridge

estimators can be developed from a Bayesian viewpoint by assuming a suitable conjugate normal prior. Then the Bayes estimator of β with respect to squared error loss is also $\hat{\beta}_k$. Hoerl and Kennard (1970) also proposed the ridge trace method to estimate k , when k is unknown. Under squared error loss, Ighodaro and Santner (1982) obtained ridge-type estimators of the multinomial cell probabilities by using the entropy distance between the cell probability vector and prior mean. These estimators are the so-called adaptive ridge estimators which can be viewed as empirical Bayes estimators, since they use estimators of the unknown parameter of the prior. Section 5 of Chapter Two is devoted to ridge-type estimators of several Poisson means under entropy loss.

In the normal case under conjugate normal priors, the risks of proper Bayes estimators of β are unbounded as $\|\beta\| \rightarrow \infty$. A similar phenomenon happens when estimating Poisson means under entropy loss. The risk of the usual MLE and UMVUE $\hat{X}$ is infinite except when $\beta = 0$, but the risks of generalized Bayes estimators are bounded. In order to retain the good risk performance of Bayes estimators of β around the prior mean and bound the risk when β is far from the prior mean, we construct compromise estimators between generalized Bayes and Bayes estimators using the "Limited

Translation Rules" proposed by Efron and Morris (1971). The risk and Bayes risk performance of these compromise estimators will be discussed in Chapter Three. Also, the "Relative Saving Loss"(RSL) from Efron and Morris (1973) will be used to compute the proportion of Bayes risk improvement over the rival generalized Bayes estimator that is sacrificed by the use of a compromise estimator instead of a Bayes estimator when the prior is true.

CHAPTER TWO
ADMISSIBLE, BAYES, HIERARCHICAL BAYES
AND RIDGE ESTIMATORS

2.1 Introduction

Let $X_1, \dots, X_p$ be p independent Poisson variables with respective means $\theta_1, \dots, \theta_p$. We write $\mathbf{X} = (X_1, \dots, X_p)$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$. Consider the problem of estimating $\boldsymbol{\theta}$ when the loss is of the form

$$L_0(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^p \theta_i^{-m_i} (\theta_i - a_i)^2, \quad (2.1.1)$$

where m_i 's are known constants. For $p=1$, admissibility of the usual (MLE, UMVUE etc.) estimator $\bar{X}$ of θ follows from a more general result of Karlin (1958) or Brown and Hwang (1982) for every m_1 (other proofs are available in Girshick and Savage (1951) or Hodges and Lehmann (1951)). Estimation of $\boldsymbol{\theta}$ in higher dimension has received considerable attention in recent years beginning with the pioneering work of Clevenson and Zidek (1975).

Clevenson and Zidek showed that when $m_1 = \dots = m_p = 1$, $\bar{X}$ was an inadmissible estimator of $\boldsymbol{\theta}$ for $p \geq 2$. Peng (1975) considered the case $m_1 = \dots = m_p = 0$, and proved the inadmissibility of $\bar{X}$ for $p \geq 3$, and its admissibility for

$p = 2$. In more recent publications, estimation of $\underline{\theta}$ is considered for general $m_1, \dots, m_p$ (see Ghosh, Hwang and Tsui (1983) for a unified treatment).

This chapter considers instead the loss as the entropy distance (or the Kullback-Leibler information number) between two distributions of p independent Poisson variables. Recall that in (1.3.1), the loss is defined as

$$L(\underline{\theta}, \underline{a}) = \sum_{i=1}^p (a_i - \theta_i - \theta_i \log \frac{a_i}{\theta_i}) . \quad (2.1.2)$$

In the above, we adopt the convention of interpreting $0/0$ as 1 and $0 \log 0$ as 0.

If the parameter space Θ of $\underline{\theta}$ is taken as $[0, \infty)^p - \{\underline{0}\}$, where $\underline{0}$, is the p -component vector with all elements equal to zeros, then under the loss (2.1.2), $\underline{X}$ has infinite risk for all $\underline{\theta} \in \Theta$. Accordingly, it is dominated by every estimator which has finite risk for at least one $\underline{\theta} \in \Theta$.

However, if the point $\underline{0} = (0, \dots, 0)$ is included in the parameter space, i.e. $\Theta = [0, \infty)^p$, then the risk of $\underline{X}$ at $\underline{\theta} = \underline{0}$ is zero, while otherwise it is $+\infty$. Also, then, $\underline{X}$ cannot be dominated by any estimator $\underline{\hat{X}}(\underline{X})$. This is because if $\underline{\hat{X}}(\underline{X})$ dominates $\underline{X}$, comparing the risks $R(\underline{\theta}, \underline{\hat{X}}(\underline{X}))$ and $R(\underline{\theta}, \underline{X})$ at $\underline{\theta} = \underline{0}$, it follows that $\underline{\hat{X}}(\underline{0}) = \underline{0}$. Accordingly, for every $\underline{\theta} \neq \underline{0}$, $R(\underline{\theta}, \underline{\hat{X}}) = +\infty$. This is clearly an example of "tyranny of the boundary of the parameter space" (see Johnson (1971))

on example 3 of page 1584).

Another way of generating estimators is to use certain priors (proper or improper) and obtain the resulting generalized Bayes estimators under the loss (2.1.2). First, note that for any prior, the generalized Bayes estimator of $\underline{\theta}$ under the loss (2.1.2) is the same as its generalized Bayes estimator under squared error loss, and is given by $E(\underline{\theta}|\underline{X})$. Thus, use of independent gamma (α_i, k_i) priors

$$g_{\underline{\alpha}, \underline{k}}(\underline{\theta}) = \prod_{i=1}^p \{ \exp(-\alpha_i \theta_i) \theta_i^{k_i} / \Gamma(k_i) \}, \quad (2.1.3)$$

$\alpha_i > 0$, $k_i > 0$ for all $i = 1, \dots, p$, leads to the proper Bayes estimators

$$\underline{e}_B(\underline{X}) = ((1 + \alpha_1)^{-1}(X_1 + k_1), \dots, (1 + \alpha_p)^{-1}(X_p + k_p)) \quad (2.1.4)$$

of $\underline{\theta}$, and these estimators have finite Bayes risks (expected loss over the joint distribution of the parameters and the samples). The class of priors described in (2.1.3) can be extended to include improper priors as well, where some or all of the α_i 's and k_i 's are allowed to be zeroes. The resulting generalized Bayes estimators are obtained as appropriate pointwise limits of the estimators derived in (2.1.4).

The above estimators are members of a more general class of estimators of the form $\underline{X}\underline{C} + \underline{b}$, where $\underline{C}$ is a

diagonal matrix of constants, and $\mathbf{b}$ is a vector of constants. In Section 2.2 of this chapter we characterize the admissible subclass within this general class of estimators. A fortiori, this will characterize the admissibility of generalized Bayes estimators under independent gamma priors or their limits. Cohen (1966) characterized admissible linear estimators of the multivariate normal mean, while Brown and Farrell (1985) characterized admissible linear estimators of Poisson means under the loss (2.1.1) with $m_1 = \dots = m_p = 1$ or 0.

There are several important consequences of the above results. So far, we have discussed only the estimation of θ . Suppose, instead, we are interested in estimating $\mathbf{h}(\theta) = (h_1(\theta_1), \dots, h_p(\theta_p))$, where each h_i is a strictly monotone increasing function on $[0, \infty)$. Then, writing the corresponding entropy loss $L(h(\theta), \mathbf{w})$ as $L(\theta, \mathbf{a})$ where $\mathbf{a} = \mathbf{h}^{-1}(\mathbf{w}) = (h_1^{-1}(w_1), \dots, h_p^{-1}(w_p))$, and L is defined in (2.1.2), it follows that the results of Sections 2.2 and 2.3 are equally applicable to the case when one estimates $\mathbf{h}(\theta)$ instead of θ . In particular, it will follow from the results of Section 2.2 that the square root statistic $((X_1 + \frac{3}{8})^{1/2}, \dots, (X_p + \frac{3}{8})^{1/2})$ of Anscombe (1948) is an admissible estimator of $(\theta_1^{1/2}, \dots, \theta_p^{1/2})$ for $p = 1, 2$, but is inadmissible for $p \geq 3$. The estimator $(\log(X_1 + 1), \dots,$

$\log(X_p+1)$) of $(\log \theta_1, \dots, \log \theta_p)$ (see Haldane (1956) or Anscombe (1956)) is admissible for $p = 1$, but is inadmissible for $p \geq 2$. The square root transformation of Anscombe (1948) is a useful variance stabilizing transformation. Also, the estimator of $(\log \theta_1, \dots, \log \theta_p)$ is very relevant especially in log-linear models. It is useful to know whether estimators dominating the classical ones achieve a substantial risk reduction.

Following Ghosh (1983), in Section 2.3 of this chapter we obtain certain proper Bayes estimators of ϑ using hierarchical priors. Such estimators have an interesting empirical Bayes interpretation. An additional important observation here is that a Bayes estimator of $h(\vartheta)$ under a prior ξ and the loss (2.1.2) is given by $h(\hat{\vartheta}_{\xi}(\tilde{X}))$ where $\hat{\vartheta}_{\xi}(\tilde{X})$ is the corresponding Bayes estimator of ϑ .

In Section 2.4, using Monte Carlo simulations, we compare the risk performances of $\tilde{X} + \underline{1}$ and the estimators dominating them. In the above, $\underline{1}$ is a p -component row vector with all elements equal to 1.

Recently, Dey and Srinivasan (1986) have considered estimation of ϑ under the loss $L_*(\vartheta, \hat{\vartheta})$

$$= \sum_{i=1}^p (\theta_i a_i^{-1} - \log(\theta_i a_i^{-1}) - 1).$$

This loss is quite different from the one given in (2.1.2). Also, Dey and Srinivasan (1986) have not attempted any characterization of

admissibility of linear estimators of the form $\tilde{X}\tilde{C} + \tilde{b}$.

2.2 The Admissibility Results

Let the parameter space be $\Theta = [0, \infty)^P$. Write $\tilde{C} = \text{diag}(c_1, \dots, c_p)$ and $\tilde{b} = (b_1, \dots, b_p)$. The main theorem of this section characterizes the admissibility of the estimators $\tilde{X}\tilde{C} + \tilde{b}$ under the loss (2.1.2).

Theorem 2.2.1. Under the loss (2.1.2), $\tilde{X}\tilde{C} + \tilde{b}$ is an admissible estimator of ϑ if and only if all three of

- (i) $b_i \geq 0$ and $c_i \geq 0$ for all $i \in S = \{1, \dots, p\}$,
- (ii) $0 \leq c_i \leq 1$ if $b_i > 0$ for all $i \in S$, and
- (iii) $0 \leq \sum_{\{i \in S: c_i = 1\}} b_i \leq 1$ hold.

The above characterization theorem is a consequence of the following lemmas. For all these lemmas, the loss is the one given in (2.1.2).

Lemma 2.2.1. $\tilde{X}\tilde{C} + \tilde{b}$ is an inadmissible estimator of ϑ if at least one of the diagonal elements of $\tilde{C}$ or an element of $\tilde{b}$ is negative.

Proof. Suppose c_ℓ and/or b_ℓ is negative for some $\ell(1 \leq \ell \leq p)$. Let $\hat{\xi}^0(\underline{X}) = \underline{X}\underline{C} + \underline{b}$ and let $\hat{\xi}^*(\underline{X})$ denote an estimator with all but its ℓ^{th} element equal to the corresponding element of $\hat{\xi}^0(\underline{X})$, while its ℓ^{th} element is equal to $(c_\ell X_\ell + b_\ell)^+$, where $a^+ = \max(a, 0)$. Then,

$$\begin{aligned} R(\hat{\xi}, \hat{\xi}^0) - R(\hat{\xi}, \hat{\xi}^*) \\ = E_{\theta_\ell} [L(\theta_\ell, c_\ell X_\ell + b_\ell) - L(\theta_\ell, (c_\ell X_\ell + b_\ell)^+)]. \end{aligned} \quad (2.2.1)$$

Note that $L(\theta_\ell, a_\ell)$ is a convex function of a_ℓ for every fixed θ_ℓ . Also, since c_ℓ and/or b_ℓ is negative, and X_ℓ assumes all nonnegative integer values with positive probability, $c_\ell X_\ell + b_\ell$ assumes negative values with positive probability. Hence, the right hand side of (2.2.1) exceeds zero, which implies that $\hat{\xi}^*$ dominates $\hat{\xi}^0$.

Lemma 2.2.2. $\underline{X}\underline{C} + \underline{b}$ is an inadmissible estimator of θ if there exists at least one $c_i > 1$ with the corresponding $b_i > 0$.

Proof. Suppose $c_\ell > 1$ and $b_\ell > 0$. Then, writing once again $\hat{\xi}^0(\underline{X}) = \underline{X}\underline{C} + \underline{b}$, but $\hat{\xi}^*(\underline{X})$ as an estimator with all but its ℓ^{th} element equal to the corresponding element of $\hat{\xi}^0(\underline{X})$, and the ℓ^{th} element of $\hat{\xi}^*(\underline{X})$ equal to $X_\ell + b_\ell c_\ell^{-1}$,

it follows that

$$\begin{aligned}
 & R(\tilde{\theta}, \tilde{\delta}^0) - R(\tilde{\theta}, \tilde{\delta}^*) \\
 &= E_{\theta_l} [L(\theta_l, c_l X_l + b_l) - L(\theta_l, X_l + b_l c_l^{-1})] \\
 &= (c_l \theta_l + b_l) - (\theta_l + b_l c_l^{-1}) - \theta_l E_{\theta_l} \log[(c_l X_l + b_l)/(X_l + b_l c_l^{-1})] \\
 &= (c_l - 1 - \log c_l) \theta_l + b_l (1 - c_l^{-1}) \\
 &> (c_l - 1 - \log c_l) \theta_l > 0, \\
 &\text{so that } \tilde{\delta}^* \text{ dominates } \tilde{\delta}^0.
 \end{aligned}$$

Lemma 2.2.3. $\tilde{X}C + \tilde{b}$ is an admissible estimator of θ if
 $0 < c_l < 1$ and $b_l > 0$ for every $l \in S$.

Proof. Consider the joint prior distribution which is the product of independent gamma $(\frac{1-c_l}{c_l}, \frac{b_l}{c_l})$ priors if $c_l > 0$ and priors degenerate at b_l if $c_l = 0$. The $\tilde{X}C + \tilde{b}$ is a proper Bayes estimator of θ under this prior with finite Bayes risk, and hence it is admissible.

Lemma 2.2.4. $\tilde{X}C$ is an admissible estimator of θ if $c_l > 0$
for all $l \in S$.

Proof. This is an immediate consequence of the admissibility of $\tilde{X}$ and the fact $R(\tilde{Q}, \tilde{X}\tilde{C}) = R(\tilde{Q}, \tilde{X}) = \tilde{Q}$ and $R(\tilde{Q}, \tilde{X}\tilde{C}) = R(\tilde{Q}, \tilde{X}) = +\infty$ if $\tilde{Q} \neq \tilde{Q}$.

Lemma 2.2.5. For every $p > 1$, $\tilde{\delta}^0(\tilde{X}) = (X_1 + b_1, \dots, X_p + b_p)$ is an admissible estimator of $\tilde{Q}$ under the loss (2.1.2) if $b_i > 0$ for all $i \in S$ and $\sum_{i \in S} b_i < 1$.

The proof of this lemma is technical, and is deferred to the section 2.5. Like many other proofs of admissibility, our proof employs Blyth's (1951) technique, but instead of using a sequence of conjugate gamma priors to approximate the improper prior with respect to which $\tilde{\delta}^0(\tilde{X})$ is generalized Bayes, our proof employs a different sequence of priors closely akin to Brown and Hwang (1982).

The final lemma of this section proves the inadmissibility of $\tilde{\delta}^0(\tilde{X})$ defined in Lemma 2.2.5 when $b_i > 0$ for all $i \in S$ and $\sum_{i \in S} b_i > 1$.

Lemma 2.2.6 Suppose $\tilde{\delta}^*(\tilde{X}) = \tilde{\delta}^0(\tilde{X}) + \phi(\tilde{X})$, when

$\phi(\tilde{X}) = (\phi_1(\tilde{X}) \dots \phi_p(\tilde{X}))$ with

$$\phi_i(\tilde{X}) = - \frac{c(\tilde{X})(X_i + b_i)}{\sum_{j=1}^p (X_j + d)}, \quad 1 \leq i \leq p$$

where $d > 0$, and $c(\underline{X})$ is nondecreasing in each coordinate.

We assume that

$$(I) \quad \sum_{j=1}^p b_j = B > 1;$$

$$(II) \quad 0 < c(\underline{X}) < \frac{2(B-1)pd}{pd+1+2(B-1)} = G \text{ (say).}$$

Then, $\hat{\underline{\theta}}^*(\underline{X})$ dominates $\hat{\underline{\theta}}^0(\underline{X})$ under the loss (2.1.2).

Proof. First note that using (2.1) of Hwang (1982),

$$\begin{aligned} R(\hat{\underline{\theta}}, \hat{\underline{\theta}}^*) - R(\hat{\underline{\theta}}, \hat{\underline{\theta}}^0) \\ &= \sum_{i=1}^p E_{\hat{\underline{\theta}}} \left(\phi_i(\underline{X}) - \theta_i \log \left(1 + \frac{\phi_i(\underline{X})}{X_i + b_i} \right) \right) \\ &= \sum_{i=1}^p E_{\hat{\underline{\theta}}} \left(\phi_i(\underline{X}) - X_i \log \left(1 + \frac{\phi_i(\underline{X} - \underline{e}_i)}{X_i + b_i - 1} \right) I_{[X_i > 1]} \right) \\ &\leq E_{\hat{\underline{\theta}}} (U(\underline{X})), \end{aligned} \tag{2.2.2}$$

where

$$U(\underline{X}) = \sum_{i=1}^p \left(\phi_i(\underline{X}) - X_i \log \left(1 + \frac{\phi_i(\underline{X} - \underline{e}_i)}{X_i + b_i - 1} \right) I_{[X_i > 1]} \right), \tag{2.2.3}$$

$\underline{e}_i$ being the i th unit vector. Next, observe that writing T

$$= \sum_{j=1}^p X_j \text{ and using assumption (II),}$$

$$\begin{aligned} &\left| \frac{\phi_i(\underline{X} - \underline{e}_i)}{X_i + b_i - 1} \right| I_{[X_i > 1]} \\ &\leq \frac{c(\underline{X} - \underline{e}_i)}{T + pd - 1} I_{[X_i > 1]} < \frac{G}{pd} < 1. \end{aligned} \tag{2.2.4}$$

Next, using (2.2.4) and Lemma 3.1 of Dey, Ghosh and Srinivasan (1986), it follows that

$$\begin{aligned}
 & \log \left(1 + \frac{\phi_1(\tilde{X} - \tilde{e}_1)}{\tilde{X}_1 + b_1 - 1} \right) I_{[X_1 > 1]} \\
 & > \left\{ \frac{\phi_1(\tilde{X} - \tilde{e}_1)}{\tilde{X}_1 + b_1 - 1} - \frac{3 - G(\text{pd})^{-1}}{6(1 - G(\text{pd})^{-1})} \cdot \frac{\phi_1^2(\tilde{X} - \tilde{e}_1)}{(\tilde{X}_1 + b_1 - 1)^2} \right\} I_{[X_1 > 1]} \\
 & > \left\{ -\frac{c(\tilde{X} - \tilde{e}_1)}{T + \text{pd} - 1} - 1/2 \text{pd}(\text{pd} - G)^{-1} \frac{c^2(\tilde{X} - \tilde{e}_1)}{(T + \text{pd} - 1)^2} \right\} I_{[X_1 > 1]}. \quad (2.2.5)
 \end{aligned}$$

Using the fact that $c(\tilde{X})$ is monotone in its arguments, it follows from (2.2.3) and (2.2.5) that

$U(\tilde{X})$

$$\begin{aligned}
 & < \left\{ \frac{c(\tilde{X})T}{T + \text{pd} - 1} + \frac{\text{pd}}{2(\text{pd} - G)} \cdot \frac{Tc^2(\tilde{X})}{(T + \text{pd} - 1)^2} \right\} I_{[T > 1]} \\
 & \quad - \frac{c(\tilde{X})(T + B)}{T + \text{pd}} \\
 & < \left(-\frac{c(\tilde{X})(T + B)}{T + \text{pd}} + \frac{c(\tilde{X})T}{T + \text{pd} - 1} + \frac{\text{pd}}{2(\text{pd} - G)} \cdot \frac{Tc^2(\tilde{X})}{(T + \text{pd} - 1)^2} \right) I_{[T > 1]} \\
 & < \frac{-c(\tilde{X})T}{(T + \text{pd})(T + \text{pd} - 1)} \left((B-1) - \frac{\text{pd} + 1}{2(\text{pd} - G)} c(\tilde{X}) \right) I_{[T > 1]} \\
 & = -\frac{c(\tilde{X})T (\text{pd} + 1)}{(T + \text{pd})(T + \text{pd} - 1)} (G - c(\tilde{X})) I_{[T > 1]} < 0, \quad (2.2.6)
 \end{aligned}$$

where assumption (II) is used in the last step. The lemma follows now from (2.2.2) and (2.2.6).

We are now in a position to prove Theorem 2.1. From Lemmas 2.2.1 and 2.2.2, it follows that $\hat{\delta}^0(\underline{x}) = \underline{x}\underline{C} + \underline{b}$ is inadmissible when a diagonal element of $\underline{C}$ or an element of $\underline{b}$ is negative, or a diagonal element of $\underline{C}$, say c_{ℓ} exceeds 1 with corresponding $b_{\ell} > 0$. Further, if some of the c_j 's equal 1 and some of the corresponding b_j 's exceed 1, then using Lemma 2.2.6, one can prove the inadmissibility of $\hat{\delta}^0(\underline{x})$. Hence, suppose that $\hat{\delta}^0(\underline{x})$ has components of the type $c_j x_j + b_j$ where each j belongs to one of the three sets

$$S_1 = \{j: 0 < c_j < 1, b_j > 0\},$$

$$S_2 = \{j: c_j = 1, b_j > 0 \text{ and } \sum_{j: c_j = 1} b_j < 1\} \text{ and}$$

$$S_3 = \{j: c_j > 0, b_j = 0\}.$$

If all j 's belong to either one of the three sets S_1 , S_2 and S_3 , appealing to Lemmas 2.2.3, 2.2.5 and 2.2.4, one proves the admissibility of $\hat{\delta}^0(\underline{x})$.

Suppose now every j belongs to one of the two sets S_1 and S_2 , neither set being empty. Suppose

$\hat{\delta}(\underline{x}) = (\delta_1(\underline{x}), \dots, \delta_p(\underline{x}))$ dominates $\hat{\delta}^0(\underline{x})$. Writing the component losses of $L(\underline{\theta}, \underline{a})$ as $L(\theta_j, a_j)$, one gets

$$E_{\underline{\theta}} \{ \sum_{j=1}^p L(\theta_j, \delta_j(\underline{x})) \} < E_{\underline{\theta}} \{ \sum_{j=1}^p L(\theta_j, \delta_j^0(\underline{x})) \}, \quad (2.2.7)$$

for all $\underline{\theta} \in [0, \infty)^p = \mathcal{M}$ with strict inequality for some $\underline{\theta} \in \mathcal{M}$. Integrating both sides of (2.2.7) with respect to pdf's corresponding to the priors $\xi_j(\theta_j)$'s described in

Lemma 2.2.3 for $j \in S_1$, one gets

$$\begin{aligned} & \int \sum_{j \in S_1} E_{\theta_j} \{L(\theta_j, \delta_j(\tilde{X}))\} \prod_{j \in S_1} d\xi_j(\theta_j) + \int \sum_{j \in S_2} E_{\theta_j} \{L(\theta_j, \delta_j(\tilde{X}))\} \prod_{j \in S_1} d\xi_j(\theta_j) \\ & < \int \sum_{j \in S_1} E_{\theta_j} \{L(\theta_j, \delta_j^0(\tilde{X}))\} \prod_{j \in S_1} d\xi_j(\theta_j) + \int \sum_{j \in S_2} E_{\theta_j} \{L(\theta_j, \delta_j^0(\tilde{X}))\} \prod_{j \in S_1} d\xi_j(\theta_j). \end{aligned} \quad (2.2.8)$$

But, since $\delta_j^0(\tilde{X})$ is the Bayes estimator of θ_j with finite Bayes risk for every $j \in S_1$, the first term in the right hand side of (2.2.8) is less than or equal to the first term in the left hand side of (2.2.8). Hence,

$$\int \sum_{j \in S_2} E_{\theta_j} \{L(\theta_j, \delta_j(\tilde{X}))\} \prod_{j \in S_1} d\xi_j(\theta_j) < \int \sum_{j \in S_2} E_{\theta_j} \{L(\theta_j, \delta_j^0(\tilde{X}))\} \prod_{j \in S_2} d\xi_j(\theta_j) \quad (2.2.9)$$

Let $\underline{\theta}_+$ denote the vector with its elements equal to θ_j ($j \in S_2$). Both the left hand side and right hand side of (2.2.9) depend on $\underline{\theta}$ only through $\underline{\theta}_+$. Also, the right hand side of (2.2.9) involves only the elements of $\underline{X}_+$, the vector with its elements X_j ($j \in S_3$). The left hand side of (2.2.9) involves the elements of $\underline{X}$, where the elements of $\underline{X}$ are independent Poissons with parameters equal to the corresponding elements of $\underline{\theta}_+$, while the remaining elements have certain negative binomial distributions depending on the b_j 's and c_j 's or are Poissons with parameters b_j 's ($j \in S_1$). Using convexity of the loss, and the sufficiency of $\underline{X}_+$ for $\underline{\theta}_+$, it follows that there exists $\delta^1(\underline{X}_+)$ such that

$$\begin{aligned}
E_{\theta_+} \left\{ \sum_{j \in S_2} L(\theta_j, \delta_j^1(\underline{X}_+)) \right\} &< \int \sum_{j \in S_2} E_{\theta_+} \{ L(\theta_j, \delta_j(\underline{X})) \} \prod_{j \in S_1} d\xi_j(\theta_j) \\
&< E_{\theta_+} \left\{ \sum_{j \in S_2} L(\theta_j, \delta_j^0(\underline{X}_+)) \right\}. \quad (2.2.10)
\end{aligned}$$

But using Lemma 2.2.5, $\delta^0(\underline{X}_+)$ is an admissible estimator of θ_+ within the class of all estimators depending only on $\underline{X}_+$.

Thus, using convexity of the loss, one must have

$\delta^1(\underline{X}_+) = \delta^0(\underline{X}_+)$ a.e. P_{θ_+} , because otherwise $\frac{1}{2}(\delta^1(\underline{X}_+) + \delta^0(\underline{X}_+))$ dominates $\delta^0(\underline{X}_+)$. Hence, one must have equality in (2.2.10), and accordingly in (2.2.9). Hence, from (2.2.8), (2.2.9) and the unique Bayesness of $\delta_j^0(\underline{X})$ with respect to the prior ξ_j , one gets

$$E_{\theta_+} \left\{ \sum_{j \in S_1} L(\theta_j, \delta_j(\underline{X})) \right\} = E_{\theta_+} \left\{ \sum_{j \in S_1} L(\theta_j, \delta_j^0(\underline{X})) \right\}. \quad (2.2.11)$$

Hence, from (2.2.8) and (2.2.11), one must have

$$E_{\theta_+} \left\{ \sum_{j \in S_2} L(\theta_j, \delta_j(\underline{X})) \right\} < E_{\theta_+} \left\{ \sum_{j \in S_2} L(\theta_j, \delta_j^0(\underline{X})) \right\}, \quad (2.2.12)$$

with strict inequality for some $\theta_+ \in \Theta$. Arguing as before with the sufficiency of $\underline{X}_+$ for θ_+ , convexity of the loss, and using Lemma 2.2.5, one finds that (2.2.12) leads to a contradiction. Thus $\delta^0(\underline{X})$ is admissible when all $j \in S_1 \cup S_2$.

It remains only to consider the case when $j \in S_3$ or j

$\in S_1 \cup S_2$, where neither S_3 nor $S_1 \cup S_2$ is empty.

Suppose $\tilde{\xi}(\underline{X})$ dominates $\tilde{\xi}^0(\underline{X})$, i.e.,

$$\begin{aligned} R(\underline{\theta}, \tilde{\xi}) &= E_{\underline{\theta}} \{ \sum_{j=1}^P L(\theta_j, \delta_j(\underline{X})) \} \\ &< E_{\underline{\theta}} \{ \sum_{j=1}^P L(\theta_j, \delta_j^0(\underline{X})) \} = R(\underline{\theta}, \tilde{\xi}^0) \end{aligned} \quad (2.2.13)$$

for all $\underline{\theta} \in \Theta$ with strict inequality for some $\underline{\theta} \in \Theta$.

Without any loss of generality, assume that $\{1, 2, \dots, p_1\} \in S_3$

and $\{p_1 + 1, \dots, p\} \in S_1 \cup S_2$ for some $p_1 \in S$. Then,

$$\begin{aligned} R(\underline{\theta}, \tilde{\xi}^0) &= R(\underline{\theta}, \tilde{\xi}^*) \text{ for all } \underline{\theta} \in \Theta \text{ with } \tilde{\xi}^*(\underline{X}) = \\ &= (0, \dots, 0, \delta_{p_1+1}^0(\underline{X}), \dots, \delta_p^0(\underline{X})). \end{aligned}$$

Write $\underline{\theta}^* = (0, \dots, 0, \theta_{p_1+1}, \dots, \theta_p)$. Then, it follows from (2.2.13) that

$$\begin{aligned} R(\underline{\theta}^*, \tilde{\xi}) &= E_{\underline{\theta}^*} \{ \sum_{j=1}^{p_1} L(0, \delta_j(\underline{X})) + \sum_{j=p_1+1}^P L(\theta_j, \delta_j(\underline{X})) \} \\ &< R(\underline{\theta}^*, \tilde{\xi}^*) = E_{\underline{\theta}^*} \{ \sum_{j=p_1+1}^P L(\theta_j, \delta_j^0(\underline{X})) \}, \end{aligned} \quad (2.2.14)$$

which implies that

$$E_{\underline{\theta}^*} \{ \sum_{j=p_1+1}^P L(\theta_j, \delta_j(\underline{X})) \} < E_{\underline{\theta}^*} \{ \sum_{j=p_1+1}^P L(\theta_j, \delta_0^j(\underline{X})) \} \quad (2.2.15)$$

Write $\underline{X}_0 = (X_{p_1+1}, \dots, X_p)$ and $\underline{\theta}_0 = (\theta_{p_1+1}, \dots, \theta_p)$.

Arguments similar as before now lead to the existence of

some $\tilde{\xi}^2(\underline{X}_0)$ such that from (2.2.15), one gets

$$E_{\varrho_0} \{ \Sigma_{j=p_1+1}^p L(\theta_j, \delta_j^2(X_0)) \} < E_{\varrho^*} \{ \Sigma_{j=p_1+1}^p L(\theta_j, \delta_j(X)) \} \\ < E_{\varrho_0} \{ \Sigma_{j=p_1+1}^p L(\theta_j, \delta_j^0(X)) \} \quad (2.2.16)$$

Since $\delta^0(X_0)$ is an admissible estimator of ϱ_0 in the class of all estimators depending only on X_0 , one must have $\delta^2(X) = \delta^0(X_0)$ a.e. P_{ϱ_0} so that one must have equality everywhere in (2.2.16), and hence, in (2.2.15). Hence, from (2.2.14)

$$E_{\varrho_0} \{ \Sigma_{j=1}^{p_1} L(0, \delta_j(X)) \} < 0,$$

so that $\delta_j(X) = 0$ a.e. for $j = 1, \dots, p_1$. Hence,

$E_{\varrho} \{ \Sigma_{j=1}^{p_1} L(\theta_j, \delta_j(X)) \} = E_{\varrho} \{ \Sigma_{j=1}^{p_1} L(\theta_j, \delta_j^0(X)) \}$ for all ϱ , each side being 0 or $+\infty$. In the event, the expression is 0; we must have

$$E_{\varrho} \{ \Sigma_{j=p_1+1}^p L(\theta_j, \delta_j(X)) \} < E_{\varrho} \{ \Sigma_{j=p_1+1}^p L(\theta_j, \delta_j^0(X)) \} \quad (2.2.17)$$

for all ϱ with strict inequality for some ϱ . Once again, use of the sufficiency of X_0 for ϱ^0 , and convexity of the loss leads to a contradiction statement from (2.2.17) since $(\delta_{p_1+1}^0(X), \dots, \delta_p^0(X))$ is an admissible estimator of ϱ^0 as proved earlier. The proof of the theorem is now complete.

2.3 Hierarchical Bayes Estimators

In this section, we develop certain Bayes estimators using hierarchical priors of Ghosh (1983). According to this formulation, conditional on $U = u (0 < u < 1)$, the θ_i 's have independent gamma $(\frac{u}{1-u}, k_i)$ prior distributions, while u has a beta (m, n) prior distribution. Then, using the results of Ghosh (1983), the Bayes estimator of ϱ is given by $\hat{\varrho}^H(\underline{X})$, where

$$\delta_j^H(\underline{X}) = \frac{n+T}{k+m+n+T} (X_j + k_j) = (1 - \frac{k+m}{k+m+n+T})(X_j + k_j), \quad (2.3.1)$$

for $j = 1, \dots, p$, where $k = \sum_{j=1}^p k_j$ and $T = \sum_{j=1}^p X_j$. These estimators have an interesting empirical Bayes interpretation. Suppose we consider only the first stage prior as described above. Then the Bayes estimator of ϱ is $(1-u)(\underline{X} + \underline{k})$, where $\underline{k} = (k_1, \dots, k_p)$. In an empirical Bayes framework, u is unknown, and is estimated from the marginal distribution of the X_i 's. It is easy to find that marginally X_i 's are independently distributed negative binomial random variables with probability functions

$$P(X_i = x_i) = \frac{\Gamma(x_i + k_i)}{x_i! \Gamma(k_i)} u^{k_i} (1-u)^{x_i}, \quad (2.3.2)$$

$x_i = 0, 1, \dots, k_i > 0$ ($i=1, \dots, p$). Then, marginally $T = \sum_{j=1}^p X_j$ is sufficient for u , having a negative binomial distribution of the form

$$P(T=t) = \frac{\Gamma(t+k)}{t! \Gamma(k)} u^k (1-u)^t. \quad (2.3.3)$$

Estimating u from the marginal distribution of T , one gets $(1-d(T))(\underline{X} + \underline{k})$ as an empirical Bayes estimator of $\underline{\theta}$. The estimator given in (2.3.1) is one such estimator with $d(T) = (k+m)/(k+m+n+T)$. If one identifies k_i with b_i in the definition of $\phi_i(\underline{X})$ in Lemma 2.2.6 and takes $c(\underline{X}) = k+m$ and $pd = p+m+n$, then, using Lemma 2.2.6, the estimator $\hat{\theta}^H(\underline{X})$ dominates the generalized Bayes estimator of $\underline{\theta}$ provided $k = \sum_{i=1}^p k_i > 1$ and $k+m < 2(k-1)(k+m+n)/(k+m+n+1+2(k-1))$. This is equivalent to $n/(k+m+n+1) > (k+m)/(2(k-1))$. This is impossible to achieve for every given m , n and p for example when $m > k-2$. However, if $m < k-2$ ($k \geq 3$), and n is very large compared to $k+m+1$, it is possible to construct a proper Bayes estimator which dominates $\underline{X}+\underline{k}$. Thus, for $k \geq 3$, it is possible to construct proper Bayes estimators using hierarchical priors which dominate $\underline{X}+\underline{k}$. One merit of constructing such estimators is that they are typically admissible.

2.4 Monte Carlo Simulations

The estimator $\underline{X}+\underline{1}$ is of special interest as discussed in the introduction. In Section 2.2, we have shown that

this estimator is inadmissible for $p > 2$. In this section, using Monte Carlo simulations, we compare the risk performances of certain estimations given in Lemma 2.2.6 which dominate $\tilde{X}+1$.

With this end, for different p , first take certain ranges of $\theta_1, \dots, \theta_p$ values. Within such ranges $\theta_1, \dots, \theta_p$ are chosen by the IMSL routine GGUBS which is the random generator of uniform distributions. Compute the risks of $\hat{\delta}^0(\tilde{X}) = \tilde{X}+1$ and $\hat{\delta}^*(\tilde{X})$, where $\hat{\delta}^*(\tilde{X})$ is given in Lemma 2.2.6 with $b_1 = \dots = b_p = 1$. We also use $c(\tilde{X}) = 2pd(p-1)/(pd+1+2p-2)$, $d = 1, 2, 3, 5, 10$, $p = 2, 5$, under the loss (2.1.2). Now, $X_1, \dots, X_p$ are generated by the IMSL routine GGPON such that X_i has a Poisson distribution with parameter θ_i , $i = 1, \dots, p$. Then, the losses $L(\hat{\delta}, \hat{\delta}^0)$ and $L(\hat{\delta}, \hat{\delta}^*)$ are calculated. The final step of the simulation procedure is repeated 1000 times to compute the simulated versions of $R(\hat{\delta}, \hat{\delta}^0)$ and $R(\hat{\delta}, \hat{\delta}^*)$ as the averages of $L(\hat{\delta}, \hat{\delta}^0)$ and $L(\hat{\delta}, \hat{\delta}^*)$ respectively. From the above, the percentage risk reduction is computed as $100[R(\hat{\delta}, \hat{\delta}^0) - R(\hat{\delta}, \hat{\delta}^*)]/R(\hat{\delta}, \hat{\delta}^0)$. The results are reported in Table 2.1. We also do similar calculations with $c(\tilde{X}) = pd(p-1)/(pd+1+2p-2)$, and the results are reported in Table 2.2.

Table 2.1

Percentage of risk reduction for $\hat{\delta}^*$ over $\hat{\delta}^0$

$$[c(\hat{\delta}) = 2pd(p-1)/(pd+1+2p-2)]$$

range of the parameter θ_i	d						
	p		1	2	3	5	10
(0,4)	2		20.03	23.86	23.82	21.23	15.21
(4,8)			3.67	4.91	5.77	6.70	6.76
(8,12)			1.99	2.39	2.76	3.39	4.08
(12,16)			1.83	2.16	2.39	2.78	3.36
(0,20)			3.00	3.77	4.32	5.08	5.58
(0,4)	5		28.54	31.98	32.03	29.37	21.99
(4,8)			9.07	10.32	10.86	11.45	11.14
(8,12)			5.36	5.98	6.17	6.53	6.99
(12,16)			3.31	3.49	3.45	3.57	4.13
(0,20)			5.55	6.21	6.42	6.78	7.20

Table 2.2

Percentage of risk reduction for $\hat{\xi}^*$ over $\hat{\xi}^0$

$$(c(\tilde{X}) = pd(p-1)/(pd+1+2p-2))$$

range of the parameter θ_i	d						
	p		1	2	3	5	10
(0,4)	2		11.86	13.55	13.16	11.37	7.90
(4,8)			2.53	3.47	3.93	4.22	3.86
(8,12)			1.43	1.93	2.21	2.50	2.59
(12,16)			1.22	1.61	1.83	2.06	2.20
(0,20)			2.00	2.70	3.07	3.40	3.35
(0,4)	5		17.30	19.44	18.96	16.61	11.79
(4,8)			5.82	7.29	7.78	7.85	6.84
(8,12)			3.46	4.42	4.81	5.03	4.78
(12,16)			2.22	2.84	3.09	3.29	3.30
(0,20)			3.56	4.55	4.94	5.17	4.88

The above simulation results are instructive in many ways. First consider the findings within a given table. Note that for a given p and d , the maximum risk reduction takes place when all the θ_i 's are in the range $(0,4)$. This is because the estimators $\hat{\xi}^*(\underline{x})$ shrink $\hat{\xi}^0(\underline{x})$ toward zero, so that the maximum risk reduction is expected when all the θ_i values are near the origin. Again for a given range of θ values, and a given d , the percentage risk reduction increases with p . Finally, for a given p , and a given range of θ values, the findings indicate that the percentage risk reduction is a concave function of d . Also, a comparison between Tables 2.4.1 and 2.4.2 reveals that percentage risk reductions when $c(\underline{x}) = 2pd(p-1)/(pd+1+2(p-1))$ is much more significant than when $c(\underline{x}) = pd(p-1)/(pd+1+2(p-1))$ for a given p and d , and a given range of θ values.

2.5 Ridge Estimators of Several Independent Poisson Means

In the usual normal theory analysis, the Bayes estimators can be viewed as ridge estimators. A similar phenomenon occurs in the Poisson case under entropy loss when one defines the distance between two points as the entropy distance rather than the Euclidean norm. In the

normal case, the entropy distance matches perfectly the Euclidean norm.

Consider the independent gamma priors given in (2.1.3) with $\alpha_1 = \dots = \alpha_p$. Then the gamma (α, k_i) prior has mean $\lambda_i = k_i/\alpha$ and the Bayes estimator of $\underline{\theta}$ given in (2.1.4) reduces to $\hat{\underline{\theta}}^\alpha(\underline{X}) = (1+\alpha)^{-1}\underline{X} + \alpha(1+\alpha)^{-1}\underline{\lambda}$, where $\underline{\lambda} = (\lambda_1, \dots, \lambda_p)$. Now suppose that $\underline{\lambda}$ is regarded as fixed. Define the entropy distance of $\underline{\theta}$ to $\underline{\lambda}$ as

$$\|\underline{\theta}\|_{\underline{\lambda}} = \sum_{i=1}^p (\theta_i - \lambda_i \log(\theta_i/\lambda_i) - \lambda_i);$$

then the following three propositions hold which show that the Bayes estimator $\hat{\underline{\theta}}^\alpha(\underline{X})$ satisfies three basic properties of ridge estimators as advocated by Hoerl and Kennard (1970).

Proposition 1. The length $L(\alpha) = \|\hat{\underline{\theta}}^\alpha(\underline{X})\|_{\underline{\lambda}}$ is a continuous decreasing function of α on $[0, \infty)$ such that $L(\alpha) \rightarrow 0$ as $\alpha \rightarrow \infty$.

Proof. Let $1/(1+\alpha) = a$. Then writing $L(\alpha) = L_0(a)$,

$$L(a) = \sum_{i=1}^p \{ (aX_i + (1-a)\lambda_i) - \lambda_i \log\left(\frac{aX_i + (1-a)\lambda_i}{\lambda_i}\right) - \lambda_i \};$$

then,

$$\begin{aligned} \frac{\partial L_0(a)}{\partial a} &= \sum_{i=1}^p \left((X_i - \lambda_i) - \frac{\lambda_i (X_i - \lambda_i)}{aX_i + (1-a)\lambda_i} \right) \\ &= \sum_{i=1}^p \left(\frac{a(X_i - \lambda_i)^2}{aX_i + (1-a)\lambda_i} \right) > 0. \end{aligned}$$

Hence,

$L_0(a) \uparrow$ as $a \uparrow$, i.e. $L(\alpha) \uparrow$ as $\alpha \uparrow$, since a is $\uparrow$ in α .

Proposition 2. For any $0 < \alpha < \infty$, $\hat{\xi}^\alpha(\underline{X})$ maximizes

$\psi(\underline{\theta}) = \sum_1^p (X_i \log \theta_i - \theta_i)$, $\psi(\underline{\theta})$ being essentially the logarithm of the likelihood among all

$$\underline{\theta} \in B_\alpha = \{ \underline{\theta} : \|\underline{\theta}\|_{\hat{\lambda}} < \|\hat{\xi}^\alpha(\underline{X})\|_{\hat{\lambda}} \}.$$

Proof.

(1) $\alpha = 0$, $\hat{\xi}^\alpha(\underline{X}) = \underline{X}$ is MLE of $\underline{\theta}$.

(2) $\alpha = \infty$, $\|\hat{\xi}^\alpha(\underline{X})\|_{\hat{\lambda}} = \|\hat{\lambda}\|_{\hat{\lambda}} = 0$. Then the result is trivial.

(3) Consider $0 < \alpha < \infty$. Note that for $\underline{\theta} \in B_\alpha$,

$$\begin{aligned} & \frac{1}{1+\alpha} \left\{ - \sum_{i=1}^p \theta_i + \sum_{i=1}^p (X_i \log \theta_i) \right\} \\ & < \frac{1}{1+\alpha} \left\{ - \sum_{i=1}^p \theta_i + \sum_{i=1}^p (X_i \log \theta_i) \right\} + \\ & \quad \frac{\alpha}{\alpha+1} \left\{ \|\hat{\xi}^\alpha(\underline{X})\|_{\hat{\lambda}} - \|\underline{\theta}\|_{\hat{\lambda}} \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^p \left\{ -\frac{\theta_i}{\alpha+1} + \frac{1}{\alpha+1} X_i \log \theta_i \right\} + \\
&\quad \sum_{i=1}^p \frac{\alpha}{\alpha+1} \left(\delta_i^\alpha(\tilde{X}) - \theta_i + \lambda_i \log \left(\frac{\theta_i}{\delta_i^\alpha(\tilde{X})} \right) \right) \\
&= \sum_{i=1}^p f_i(\theta_i) \quad (\text{say}).
\end{aligned}$$

The expression in brackets is concave in θ_i .

This is because

$$\frac{\partial f_i(\theta_i)}{\partial \theta_i} = -1 + \frac{X_i}{(\alpha+1)\theta_i} + \frac{\alpha\lambda_i}{(1+\alpha)\theta_i}, \quad (2.6.1)$$

$$\text{and} \quad \frac{\partial^2 f_i(\theta_i)}{\partial \theta_i^2} = - \left\{ \frac{X_i}{(1+\alpha)\theta_i^2} + \frac{\alpha\lambda_i}{(1+\alpha)\theta_i^2} \right\} < 0.$$

From (2.6.1), we find that $\sum_{i=1}^p f_i(\theta_i)$ is maximized at

$\theta_i = \delta_i^\alpha(\tilde{X}) \quad (i = 1, \dots, p)$. Putting $\theta_i = \delta_i^\alpha(\tilde{X})$ in $f_i(\theta_i)$, one gets,

$$\begin{aligned}
&\frac{1}{\alpha+1} \sum_{i=1}^p (X_i \log \theta_i - \theta_i) \\
&< \frac{1}{\alpha+1} \sum_{i=1}^p \{ (X_i \log \delta_i^\alpha(\tilde{X}) - \delta_i^\alpha(\tilde{X})) \}
\end{aligned}$$

which proves Proposition 2.

Proposition 3. For every $\alpha > 0$, for any ϱ with at least one non-zero component $E \left\| \left| \delta^\alpha(\tilde{X}) \right| \right\|_{\varrho} < E \left\| \left| \tilde{X} \right| \right\|_{\varrho}$.

Proof. $E \left\| \frac{X}{\lambda} \right\|_{\vartheta} = R(\vartheta, \tilde{X}) = \infty$ under the loss (2.1.2).

Remark 2.5.1 Under the loss (2.1.1), the Bayes estimator of ϑ with respect to the prior (2.1.3) is

$$\delta_{\tilde{m}}(\tilde{X}) = (\delta_{m_1}(\tilde{X}), \dots, \delta_{m_p}(\tilde{X})) \text{ where}$$

$$\delta_{m_i}(\tilde{X}) = \frac{X_i + k_i - m_i}{1 + \alpha_i}, \quad i = 1, \dots, p \quad (k_i > m_i). \quad \text{Let } \tilde{\lambda} =$$

$$\left(\frac{k_1 - m_1}{\alpha}, \dots, \frac{k_p - m_p}{\alpha} \right) = (\lambda_1, \dots, \lambda_p). \quad \text{Define the entropy}$$

distance from ϑ to $\tilde{\lambda}$

$$\text{as } \left\| \vartheta \right\|_{\tilde{\lambda}} = \sum_{i=1}^p \left(\theta_i - \lambda_i \log \left(\frac{\theta_i}{\lambda_i} \right) - \lambda_i \right). \quad \text{In this case, also,}$$

the above propositions hold.

For unknown α , these propositions suggest an adaptive ridge method for estimating α . One possible approach is to use the method, from Southerland, Fienberg and Holland (1974), which finds a α value such that the risk $R(\vartheta, (X + \alpha \tilde{\lambda}) / (1 + \alpha))$ is minimized. In this set up, the α value involves $\theta = \sum_{i=1}^p \theta_i$, so that one has to estimate θ from data. In practice, a Monte Carlo study can be employed to assess the behavior of estimators for this method.

2.6 Proof of Lemma 2.2.5

Proof. Under the prior $\pi(\underline{\theta}) = \left(\prod_{j=1}^p \theta_j^{b_j-1} \right) g(\underline{\theta})$

where $b_j > 0$ for all j , and the loss (2.1.2), the Bayes estimator of $\underline{\theta}$ is

$\underline{\delta}^g(\underline{X}) = (\delta_1^g(\underline{X}), \dots, \delta_p^g(\underline{X}))$, where

$$\begin{aligned} \delta_i^g(x) &= \left\{ \int_0^\infty \dots \int_0^\infty \theta_i \exp\left(-\sum_{j=1}^p \theta_j\right) \prod_{j=1}^p \theta_j^{x_j+b_j-1} g(\underline{\theta}) d\theta_1 \dots d\theta_p \right\} \\ &\div \left\{ \int_0^\infty \dots \int_0^\infty \exp\left(-\sum_{j=1}^p \theta_j\right) \prod_{j=1}^p \theta_j^{x_j+b_j-1} g(\underline{\theta}) d\theta_1, \dots, d\theta_p \right\}. \end{aligned} \quad (2.6.1)$$

Define $I_{\underline{X}}(g)$ as the denominator given in (2.6.1). Then, integration by parts leads to

$$\delta_i^g(\underline{x}) = x_i + b_i + \{I_{\underline{x}+\underline{e}_i}(\nabla_i g) / I_{\underline{x}}(g)\}, \quad (2.6.2)$$

where $\nabla_i g = \frac{\partial g}{\partial \theta_i}$ and $\underline{e}_i$ is the i th unit vector. Take $g(\underline{\theta}) =$

1 and $g_n(\underline{\theta}) = g(\underline{\theta}) h_n^2(\underline{\theta})$, where $h_n^2(\underline{\theta})$ will be defined

later. Then $I_{\underline{x}+\underline{e}_i}(\nabla_i g) = 0$ and one gets from (2.6.2)

$$\delta_i^g(\underline{x}) - \delta_i^{g_n}(\underline{x}) = - \frac{I_{\underline{x}+\underline{e}_i}(\nabla_i g_n)}{I_{\underline{x}}(g_n)}. \quad (2.6.3)$$

Hence, writing $r(\pi, \underline{\delta})$ as the Bayes risk of an estimator $\underline{\delta}$ of $\underline{\theta}$ with respect to the prior π , and

writing $\pi_n(\underline{\theta}) = \prod_{j=1}^p \theta_j^{b_j-1} g_n(\underline{\theta})$, one gets

$$\begin{aligned}
& r(\pi_n, \hat{\xi}_n^g) - r(\pi_n, \hat{\xi}_n^{g_n}) \\
&= \sum_{i=1}^p E \left\{ \delta_i^g(\underline{x}) - \delta_i^{g_n}(\underline{x}) - \theta_i \log \frac{\delta_i^g(\underline{x})}{\delta_i^{g_n}(\underline{x})} \right\}, \quad (2.6.4)
\end{aligned}$$

where E denotes expectation over the joint distribution of $\underline{\theta}$ and $\underline{x}$. Note that

$$\begin{aligned}
& E\{\delta_i^g(\underline{x}) - \delta_i^{g_n}(\underline{x})\} \\
&= - \sum_{x_1=0}^{\infty} \dots \sum_{x_p=0}^{\infty} \int_0^{\infty} \dots \int_0^{\infty} \exp\left(-\sum_{j=1}^p \theta_j\right) \left(\prod_{j=1}^p \theta_j^{x_j+b_j-1} / \prod_{j=1}^p x_j!\right) \theta_i \frac{\partial g_n}{\partial \theta_i} d\theta_1 \dots d\theta_p \\
&= - \int_0^{\infty} \dots \int_0^{\infty} \left(\prod_{j=1}^p \theta_j^{b_j-1}\right) \theta_i \frac{\partial g_n}{\partial \theta_i} d\theta_1 \dots d\theta_p, \quad (2.6.5)
\end{aligned}$$

and

$$\begin{aligned}
& -E\left\{\theta_i \log \frac{\delta_i^g(\underline{x})}{\delta_i^{g_n}(\underline{x})}\right\} \\
&= -E\left\{\theta_i \log \frac{x_i + b_i}{x_i + b_i + \{I_{\hat{\xi}_n^g}(\nabla_i g_n) / I_{\hat{\xi}_n^{g_n}}(g_n)\}}\right\} \\
&= E\left\{\theta_i \log \left(\frac{x_i + b_i}{x_i + b_i + \{I_{\hat{\xi}_n^g}(\nabla_i g_n) / I_{\hat{\xi}_n^{g_n}}(g_n)\}}\right)^{-1}\right\} \\
&\leq E\left\{\theta_i \frac{I_{\hat{\xi}_n^g}(\nabla_i g_n)}{(x_i + b_i) I_{\hat{\xi}_n^{g_n}}(g_n)}\right\} \\
&= \sum_{x_1=0}^{\infty} \dots \sum_{x_p=0}^{\infty} \frac{1}{\prod_{j=1}^p x_j!} \cdot \frac{I_{\hat{\xi}_n^g + \hat{\xi}_n^{g_n}}(\nabla_i g_n) I_{\hat{\xi}_n^g + \hat{\xi}_n^{g_n}}(g_n)}{(x_i + b_i) I_{\hat{\xi}_n^{g_n}}(g_n)}. \quad (2.6.6)
\end{aligned}$$

Next noting that

$I_{\tilde{x}+e_i}(g_n) = (x_i + b_i) I_{\tilde{x}}(g_n) + I_{\tilde{x}+e_i}(\nabla_i g_n)$, one gets from (2.6.6),

$$\begin{aligned}
 & -E\left\{\theta_i \log \frac{\delta_i^g(\tilde{x})}{\delta_i^{g_n}(\tilde{x})}\right\} \\
 &= x_1^{\infty} \dots x_p^{\infty} \sum_{j=1}^{\infty} \frac{1}{p x_j!} \left\{ I_{\tilde{x}+e_i}(\nabla_i g_n) + \frac{I_{\tilde{x}+e_i}^2(\nabla_i g_n)}{(x_i + b_i) I_{\tilde{x}}(g_n)} \right\} \\
 &= \int_0^{\infty} \dots \int_0^{\infty} \left(\prod_{j=1}^p \theta_j^{b_j-1} \right) \theta_i \frac{\partial g_n}{\partial \theta_i} d\theta_1 \dots d\theta_p \\
 &+ x_1^{\infty} \dots x_p^{\infty} \sum_{j=1}^{\infty} \frac{1}{p x_j!} \cdot \frac{I_{\tilde{x}+e_i}^2(\nabla_i g_n)}{(x_i + b_i) I_{\tilde{x}}(g_n)} . \tag{2.6.7}
 \end{aligned}$$

Combining (2.6.4), (2.6.5) and (2.6.7), one gets

$$\begin{aligned}
 & r(\pi_n, \tilde{x}^g) - r(\pi_n, \tilde{x}^{g_n}) \\
 &< \sum_{i=1}^p x_1^{\infty} \dots x_p^{\infty} \sum_{j=1}^{\infty} \frac{1}{\left(\prod_{j=1}^p x_j! \right) (x_i + b_i)} \cdot \frac{I_{\tilde{x}+e_i}^2(\nabla_i g_n)}{I_{\tilde{x}}(g_n)} \\
 &< \sum_{i=1}^p x_1^{\infty} \dots x_p^{\infty} \sum_{j=1}^{\infty} \frac{1}{\left(\prod_{j=1}^p x_j! \right) (x_i + b_i)} \int_0^{\infty} \dots \int_0^{\infty} e^{1-\sum \theta_j} \theta_i^2 \left(\prod_{j=1}^p \theta_j^{x_j+b_j-1} \right) \left(\frac{\partial h_n}{\partial \theta_i} \right)^2 d\theta_1 \dots d\theta_p
 \end{aligned}$$

$$\leq \sum_{i=1}^p (1+b_i^{-1}) x_i \sum_{x_i=0}^{\infty} ((x_i+1)!)^{-1} \int_0^{\infty} \dots \int_0^{\infty} e^{-\theta_i x_i + 1} \theta_i (\sum_{j=1}^p \theta_j^{b_j-1}) \left(\frac{\partial h}{\partial \theta_i} \right)^2 d\theta_1 \dots d\theta_p$$

$$\leq (1 + \max_{1 \leq i \leq p} b_i^{-1}) \sum_{i=1}^p \int_0^{\infty} \dots \int_0^{\infty} \theta_i (\sum_{j=1}^p \theta_j^{b_j-1}) \left(\frac{\partial h}{\partial \theta_i} \right)^2 d\theta_1 \dots d\theta_p. \quad (2.6.8)$$

Now take

$$h_n(\theta) = 1 \quad \text{if } 0 < \mu = \sum_{i=1}^p \theta_i < 1$$

$$= 1 - \frac{\log \mu}{\log n} \quad \text{if } 1 < \mu < n$$

$$= 0, \quad \text{otherwise.} \quad (2.6.9)$$

Then, from (2.6.8), it follows that

$$r(\pi_n, \tilde{\delta}_n^g) - r(\pi_n, \tilde{\delta}_n^g)$$

$$\leq (1 + \max_{1 \leq i \leq p} b_i^{-1}) (\log n)^{-2} \int \dots \int_{1 < \mu < n} \mu^{-1} (\sum_{j=1}^p \theta_j^{b_j-1}) d\theta_1 \dots d\theta_p. \quad (2.6.10)$$

Using the transformation $\mu = \sum_{i=1}^p \theta_i, \phi_i = \theta_i / \mu$ ($i = 1, \dots, p-1$),

it follows that the integral in the right hand side of

(2.6.10) equals

$$\int \dots \int_{\substack{p-1 \\ \phi_j > 0, \sum_{j=1}^p \phi_j < 1}} \int_1^n \mu^{-1} \left(\mu^{\sum_{j=1}^p b_j - p + p - 1} \right) \prod_{j=1}^{p-1} \theta_j^{b_j-1} \left(1 - \sum_{j=1}^{p-1} \phi_j \right)^{b_p-1} d\mu d\phi_1 \dots d\phi_{p-1}$$

$$\leq \text{const} \int_1^n \mu^{\sum_{j=1}^p b_j - 2} d\mu$$

$$\leq \text{const} \text{ if } \sum_{j=1}^p b_j < 1$$

$$\leq \text{const} (\log n) \text{ if } \sum_{j=1}^p b_j = 1. \quad (2.6.11)$$

It follows from (2.6.10) and (2.6.11) that

$r(\pi_n, \hat{\xi}^g) - r(\pi_n, \hat{\xi}^{g_n}) \rightarrow 0$ as $n \rightarrow \infty$ when $\sum_{j=1}^p b_j < 1$, and using Blyth's theorem (see for example Brown and Hwang (1983));

the proof of the lemma is complete.

CHAPTER THREE COMPROMISE BETWEEN GENERALIZED BAYES AND BAYES ESTIMATORS

3.1 Introduction

We have found in the previous chapter that the Bayes estimators of multiple Poisson means $\theta_1, \dots, \theta_p$ under independent gamma priors (2.1.3) and the entropy loss (2.1.2) are given by $\delta^B(\underline{X}) = (\delta_1^B(\underline{X}), \dots, \delta_p^B(\underline{X}))$, where $\delta_i^B(\underline{X}) = (1 + \alpha_i)^{-1}(X_i + k_i)$ for all $i = 1, \dots, p$. Obviously the Bayes estimators $\delta^B(\underline{X})$ are all admissible, while the generalized Bayes estimators $(X_1 + k_1, \dots, X_p + k_p)$ with $k_i > 0$ ($i = 1, \dots, p$) are admissible if and only if $\sum_{i=1}^p k_i < 1$. However, we shall presently see that the Bayes estimators suffer from a lack of robustness against misspecified priors to which the generalized Bayes estimators are not subject.

To see this, without any loss of generality, take $p = 1$. Let X be a Poisson variable with parameter $\theta \in [0, \infty)$, and suppose θ has a $\text{gamma}(\alpha, k)$ prior. Then the Bayes estimator $\delta_{\alpha, k}$ of θ under the entropy loss (2.1.2) is $(1 + \alpha)^{-1}(X + k)$, and has the risk

$$\begin{aligned}
& R(\theta, \delta_{\alpha, k}) \\
&= E_{\theta} \{ (1 + \alpha)^{-1} (X + k) - \theta \log((X + k)/(1 + \alpha)) + \theta \log \theta - \theta \} \\
&= k(1 + \alpha)^{-1} + \theta((1 + \alpha)^{-1} - 1 - \log(1 + \alpha)^{-1}) - \theta E_{\theta} \log((X + k)\theta^{-1}) \\
&> k(1 + \alpha)^{-1} + \theta((1 + \alpha)^{-1} - 1 - \log(1 + \alpha)^{-1}) - \theta \log((\theta + k)\theta^{-1}), \quad (3.1.1)
\end{aligned}$$

by Jensen's inequality. Since $\theta \log((\theta + k)\theta^{-1}) < k$ and $(1 + \alpha)^{-1} - \log(1 + \alpha)^{-1} - 1 > 0$ for $\alpha > 0$, it follows that $R(\theta, \delta_{\alpha, k}) \rightarrow \infty$ as $\theta \rightarrow \infty$. The same is not true, however, for a generalized Bayes estimator $X + k$ which has risk bounded in θ . To see this, first note that the risk of the estimator $\delta^k(X) = X + k$ is

$$R(\theta, \delta^k) = k - \theta E_{\theta}(\log(X + k)) + \theta \log \theta, \quad (3.1.2)$$

which is continuous in θ . Further, it is immediate from (3.1.2) that

$$\lim_{\theta \rightarrow 0} R(\theta, \delta^k) = k. \quad (3.1.3)$$

Next we prove the following two lemmas which show that $\sup_{\theta} R(\theta, \delta^k)$ is bounded.

Lemma 3.1.1. Under the entropy loss (2.1.2),

$$\sup_{\theta} R(\theta, \delta^k) = k \text{ if } k \geq 1. \quad (3.1.4)$$

Proof. First note that for $k \geq 1$,

$$\begin{aligned} \theta E_{\theta} \log(X + k) &= \sum_{x=0}^{\infty} (x+1) \log(x+k) \exp(-\theta) \frac{\theta^{x+1}}{(x+1)!} \\ &= \sum_{x=1}^{\infty} x \log(x+k-1) \exp(-\theta) \frac{\theta^x}{x!} \\ &= E_{\theta}(X \log(X+k-1)). \end{aligned} \quad (3.1.5)$$

Next observe that the function $g(x) = x \log(x+k-1)$ is a convex function of x when $k \geq 1$. Hence, using Jensen's inequality and interpreting $0 \log 0$ as 0, it follows from (3.1.2) and (3.1.5) that

$$R(\theta, \delta^k) \leq k - \theta \log(\theta + k - 1) + \theta \log \theta \leq k, \quad (3.1.6)$$

since $k \geq 1$. Now, (3.1.4) follows (3.1.3) and (3.1.6).

Lemma 3.1.2. Under the entropy loss (2.1.2),

$$R(\theta, \delta^1) - R(\theta, \delta^k) \rightarrow 0 \text{ as } \theta \rightarrow \infty. \quad (3.1.7)$$

Proof. Note that

$$\begin{aligned} R(\theta, \delta^1) - R(\theta, \delta^k) &= 1 - k + \theta E_{\theta} \log\{(X+k)/(X+1)\} \\ &= 1 - k + \sum_{x=0}^{\infty} \{\log\{(x+k)/(x+1)\} (x+1) \exp(-\theta) \theta^{x+1} / (x+1)!\} \\ &= 1 - k + E_{\theta}(X \log(1 + (k-1)X^{-1}) I_{(X \geq 1)}) \end{aligned}$$

$$\longrightarrow 1 - k + k - 1 = 0, \quad (3.1.8)$$

as $\theta \rightarrow \infty$ by the Lebesgue dominated convergence theorem since $X \log(1 + (k-1)X^{-1}) I_{(X > 1)} \leq k-1$ and $X \xrightarrow{a.s.} \infty$ as $\theta \rightarrow \infty$.

From (3.1.3), (3.1.4) and (3.1.7), it follows that $R(\theta, \delta^k)$ remains bounded as $\theta \rightarrow 0$ or $\theta \rightarrow \infty$ for all $k > 0$. Since $R(\theta, \delta^k)$ is continuous in θ , it follows that $R(\theta, \delta^k)$ remains bounded for all θ .

On the other hand, it is easy to see that when $\theta = k \alpha^{-1}$, the prior mean, then

$$R(\theta, \delta^k) - R(\theta, \delta_{\alpha,k}) = k(1 - \alpha^{-1} \log(1+\alpha)) > 0, \quad (3.1.9)$$

for all $\alpha > 0$. Hence, as anticipated, for every $k > 0$ the Bayes estimator $\delta_{\alpha,k}$ dominates the corresponding generalized Bayes estimator δ^k around the prior mean. Thus, while the Bayes estimator performs very satisfactorily around the prior mean, it is highly vulnerable against misspecified priors.

The vulnerability of δ^k against wrong priors is also reflected in the following Bayesian analysis. To see this consider the prior Π_1 which is gamma (α_1, k_1) . Writing $r(\Pi_1, \delta)$ as the Bayes risk of an estimator δ of θ under the prior Π_1 , one gets

$$r(\Pi_1, \delta_{\alpha,k}) - r(\Pi_1, \delta^k)$$

$$\begin{aligned}
&= E_{\Pi_1} (R(\theta, \delta_{\alpha, k}) - R(\theta, \delta^k)) \\
&= E_{\Pi_1} \{ k(1+\alpha)^{-1} - k + \theta((1+\alpha)^{-1} - 1 - \log(1+\alpha)^{-1}) \} \\
&= -k\alpha(1+\alpha)^{-1} + k_1 \alpha_1^{-1} ((1+\alpha)^{-1} - 1 - \log(1+\alpha)^{-1}). \quad (3.1.10)
\end{aligned}$$

Obviously for fixed k and α , the right hand side of (3.1.10) is an unbounded function of $k_1 \alpha_1^{-1}$. The bigger the value of $k_1 \alpha_1^{-1}$, the bigger is the Bayes risk difference.

Our objective in this chapter is to propose certain estimators which serve as a compromise between Bayes and generalized Bayes estimators. We shall use "Limited Translation Rules" as a compromise method which was proposed by Efron and Morris (1971) in the normal case. In practice we shall restrict ourselves to $0 < k \leq 1$, since estimators of the form $X + k$ with $k > 1$ have already been proved to be inadmissible.

Define the compromise estimator of θ between $\delta_{\alpha, k}$ and $X + C$ (say δ^C) as

$$\delta_{M, \alpha, k}^C = \begin{cases} \frac{1}{1+\alpha} X + \frac{k}{1+\alpha}, & X < \frac{(1+\alpha)(M-C) + k}{\alpha} \equiv d, \\ X + C - M, & \text{otherwise,} \end{cases}$$

where $M > \max(0, C - k/(1+\alpha))$ and $[u]$ denotes the greatest integer less than or equal to u . The above estimators

$\delta_{M,\alpha,k}^C$ fixes the maximum allowable deviation, say M , from $X + C$ and use the Bayes estimator $\delta_{\alpha,k}$ subject to the constraint $X + C - (1+\alpha)^{-1}(X + k) < M$. It is also a shrinkage estimator which shrinks more towards δ^C for large values of X . Indeed, the data determine the shrinking factor.

In Section 3.2, we investigate the risk performance of $\delta_{M,\alpha,k}^C$. In Section 3.3, we shall employ the idea of relative saving loss (RSL) as introduced by Efron and Morris (1971) to compare $\delta_{M,\alpha,k}^C$ and the Bayes estimator $\delta_{\alpha,k}$ in terms of their Bayes risk performance. Some figures and tables showing the risks as well as the RSL's of limited translation rules $\delta_{M,\alpha,k}^C$ are given in Sections 3.2 and 3.3.

3.2 The Risk Performance of $\delta_{M,\alpha,k}^C$

Write $1/(1+\alpha) = a$ and $k/(1+\alpha) = b$. The $\delta_{M,\alpha,k}^C$ is defined by

$$\delta_{M,\alpha,k}^C = \begin{cases} aX + b, & X < \frac{(M - C) + b}{1 - a} \equiv d, \\ X + C - M, & X > d, \end{cases} \quad (3.2.1)$$

where $M > \max(0, C - b)$ and $[u]$ denotes the greatest integer which is less than or equal to u . The first theorem of this section shows the estimator $\delta_{M,\alpha,k}^C$ has risk bounded

for all θ .

Theorem 3.2.1. $\lim_{\theta \rightarrow \infty} R(\theta, \delta_{M,\alpha,k}^C) < 1$ and $R(\theta, \delta_{M,\alpha,k}^C)$ is bounded function of θ .

Proof. Note that

$$\begin{aligned} & R(\theta, \delta_{M,\alpha,k}^C) \\ &= E_{\theta} \{ ((aX + b) - \theta \log(aX + b)) I_{(X \leq d)} \} + \\ & \quad E_{\theta} \{ ((X + C - M) - \theta \log(X + C - M)) I_{(X > d)} \} - \theta + \theta \log \theta \\ &= A_d(\theta) + B_d(\theta) + \theta \log \theta \quad (\text{say}), \end{aligned} \tag{3.2.2}$$

where $A_d(\theta)$ is equal to the first term of the line above

(3.2.2) and $B_d(\theta)$ equal to the sum of the second term and the third term. It is clear that $A_d(\theta) \rightarrow 0$ as $\theta \rightarrow \infty$.

Also $B_d(\theta) = (1 - F_{\theta}(d)) E_{\theta}[(Y + C - M) - \theta \log(Y + C - M)] - \theta$ where Y has a Poisson distribution truncated at d and F_{θ} is a distribution function of Poisson variable with parameter θ .

Now,

$$\begin{aligned} B_d(\theta) &= (1 - F_{\theta}(d)) \left(\frac{\theta(1 - F_{\theta}(d-1))}{1 - F_{\theta}(d)} + C - M \right) - \theta \\ &\quad - (1 - F_{\theta}(d)) \{ E_{\theta}(\theta \log(Y + C - M)) \}, \end{aligned}$$

where

$$(1 - F_{\theta}(d)) \left(\frac{\theta(1 - F_{\theta}(d-1))}{1 - F_{\theta}(d)} + C - M \right) - \theta + C - M \text{ as } \theta \rightarrow \infty.$$

Since $f(x) = x \log(x + C - 1 - M)$ is a convex function when $x \geq 2(M + 1 - C)$, we consider

$$(1 - F_{\theta}(d)) E_{\theta}(\theta \log(Y + C - M))$$

$$= \sum_{x=d+1}^{2(d+2)-1} \theta \log(x + C - M) \frac{\exp(-\theta) \theta^x}{x!} \\ + \sum_{x=2(d+2)}^{\infty} \theta \log(x + C - M) \frac{\exp(-\theta) \theta^x}{x!}$$

$$= D_d(\theta) + (1 - F_{\theta}(2d+3)) E_{\theta}(Z \log(Z + C - 1 - M) I(Z \geq 2d+4)), (3.2.3)$$

where $D_d(\theta)$ is the first term of the above expression and Z has a Poisson distribution truncated at $2d + 3$. Also, it is immediate that $D_d(\theta) \rightarrow 0$ as $\theta \rightarrow \infty$. Using Jensen's inequality, the second term in (3.2.3) is greater than or equal to

$$\theta(1 - F_{\theta}(2d+2)) \log \left\{ \frac{\theta(1 - F_{\theta}(2d+2))}{(1 - F_{\theta}(2d+3))} + C - 1 - M \right\}.$$

Finally,

$$\lim_{\theta \rightarrow \infty} R(\theta, \delta_{M, \alpha, k}^C)$$

$$\leq (C - M) -$$

$$\lim_{\theta \rightarrow \infty} \{ \theta(1 - F_{\theta}(2d+2)) \log \left(\frac{\theta(1 - F_{\theta}(2d+2))}{(1 - F_{\theta}(2d+3))} + C-1-M \right) - \theta \log \theta \}$$

$$= (C - M) +$$

$$\lim_{\theta \rightarrow \infty} \{ \theta F_{\theta}(2d+2) \log \left(\frac{\theta(1 - F_{\theta}(2d+2))}{1 - F_{\theta}(2d+3)} + C - 1 - M \right) \} -$$

$$\lim_{\theta \rightarrow \infty} \theta \log \left(\frac{1 - F_{\theta}(2d+2)}{1 - F_{\theta}(2d+3)} - \frac{M + 1 - C}{\theta} \right)$$

$$= (C - M) + 0 + M + 1 - C = 1.$$

Since $\lim_{\theta \rightarrow 0} R(\theta, \delta_{M,\alpha,k}^C) = k/(1+\alpha) = b$ and $R(\theta, \delta_{M,\alpha,k}^C)$ is

a continuous function of θ , $R(\theta, \delta_{M,\alpha,k}^C)$ is bounded. The proof is complete.

Theorem 3.3.2.

$$\sup_{\theta} R(\theta, \delta_{M,\alpha,k}^C)$$

$$\leq \begin{cases} ad+b-\log b + |C-M| + (d+2)\log(d+1+C-M) + 2(d+2)\log(2d+4), & 0 < b < 1 \\ ad+b + |C-M| + 2(d+2)\log(2d+4) & , \quad b \geq 1 \end{cases}$$

Proof. The following $A_d(\theta)$ and $B_d(\theta)$ are from the proof of theorem 3.2.1. Now,

$$A_d(\theta)$$

$$= \sum_{x=0}^d ((ax+b) - \theta \log(ax+b)) \frac{\exp(-\theta) \theta^x}{x!}$$

$$= \sum_{x=0}^d (ax+b) \frac{\exp(-\theta)\theta^x}{x!} - \sum_{x=1}^{d+1} x \log(a(x-1)+b) \frac{\exp(-\theta)\theta^x}{x!}$$

$$\leq ad + b - \log b \sum_{x=1}^{d+1} \frac{\exp(-\theta)\theta^x}{x!}$$

$$\leq \begin{cases} ad + b - \log b & , \quad 0 < b < 1, \\ ad + b & , \quad b > 1, \end{cases}$$

$$B_d(\theta) + \theta \log \theta$$

$$= \left\{ \left(\sum_{x=d+1}^{\infty} (x + C - M) \frac{\exp(-\theta)\theta^x}{x!} \right) - \theta \right\} -$$

$$\left\{ \left(\sum_{x=d+1}^{\infty} \theta \log \theta (x + C - M) \frac{\exp(-\theta)\theta^x}{x!} \right) - \theta \log \theta \right\}. \quad (3.2.4)$$

$$\text{The first term in (3.2.4)} \leq |C - M| + \sum_{x=d+1}^{\infty} \frac{\exp(-\theta)\theta^x}{(x-1)!} - \theta$$

$$\leq |C - M|.$$

The second term in (3.2.4) is equal to

$$\sum_{x=d+1}^{2d+3} \theta \log(x + C - M) \frac{\exp(-\theta)\theta^x}{x!} +$$

$$\sum_{x=2(d+2)}^{\infty} \theta \log(x + C - M) \frac{\exp(-\theta)\theta^x}{x!} - \theta \log \theta,$$

$$> \begin{cases} 0 + (1 - F_{\theta}(2d+3))E_{\theta}(Z \log(Z+C-1-M)) - \theta \log \theta, & b \geq 1 \\ (d+2) \log(d+1+C-M) + (1 - F_{\theta}(2d+3))E_{\theta}(Z \log(Z+C-1-M)) - \theta \log \theta, & 0 < b < 1 \end{cases}$$

$$0 < b < 1$$

where Z is a Poisson distribution truncated at $2d + 3$.

By definition $2d + 3 + C - M > 1$. Define

$$f(x) = \begin{cases} x \log(x + C - 1 - M), & x \geq 2(d+2) \\ 2(d+2) \log(2d + 3 + C - M), & x < 2(d+2). \end{cases}$$

Then

$$(1 - F_\theta(2d+3))E_\theta[Z \log(Z + C - 1 - M)] - \theta \log \theta$$

$$= E[f(Z)] -$$

$$\sum_{x=0}^{2(d+2)} 2(d+2) \log(2(d+2) + C - 1 - M) \frac{\exp(-\theta) \theta^x}{x!} - \theta \log \theta$$

$$\geq f[E(Z)] - \theta \log \theta - 2(d+2) \log(2(d+2) + C - 1 - M)$$

$$= \begin{cases} \theta \log\left(\frac{\theta + C - 1 - M}{\theta}\right) - 2(d+2) \log(2(d+2) + C - 1 - M), & \theta \geq 2(d+2) \\ 2(d+2) \log(2(d+2) + C - 1 - M) - \theta \log \theta - 2(d+2) \log(2(d+2) + C - 1 - M), & \theta < 2(d+2) \end{cases}$$

$$\geq \begin{cases} -2(d+2) \log(2(d+2)), & \theta \geq 2(d+2) \\ -2(d+2) \log(2(d+2)), & \theta < 2(d+2). \end{cases}$$

Thus, we have

$$\sup_{\theta} R(\theta, \delta_{M, \alpha, k}^C)$$

$$\leq \begin{cases} ad + b + |C-M| + 2(d+2) \log(2d+4), & b \geq 1 \end{cases}$$

$$\begin{cases} ad + b - \log b + |C-M| + (d+2)\log(d+1+C-M) + \\ 2(d+2)\log(2d+4), \end{cases} \quad b < 1.$$

The above bounds, though explicit are not necessarily sharp. This will be revealed in Figures 3.1 and 3.2. Figure 3.1 shows the risk of $\delta_{M,\alpha,k}^C$ as a function of θ where $M = 1$ or 3 and $C = 3/8$. We plot also the risk of $\delta^{3/8}$, δ^k and $\delta_{\alpha,k}$. The parameters α and k take the value $1/3$ and 1 , respectively. We see from this figure that the limited translation rules $\delta_{M,1/3,1}^{3/8}$, with $M = 1$ and 3 , perform as well as the Bayes estimator $\delta_{1/3,1}$ around the prior mean 3 . But the risk of the Bayes estimator $\delta_{1/3,1}$ linearly increases when $\theta \geq 10$, whereas the risk of $\delta_{1,1/3,1}^{3/8}$ never does exceed $.6$ whereas the risk of $\delta_{3,1/3,1}^{3/8}$ never exceeds $.9$. Instead of $C = 3/8$, we take $C = 1$ in the figure 3.2 with the same α , k and M values. The risk performances of the limited translation rules in Figure 3.2 are similar to those in Figure 3.1. Also, from these figures it follows that the limited translation rule with $M = 1$ has in general superior risk performance compared to the limited translation rule with $M = 3$. Table 3.1 shows the values of the risks of the different estimators considered in Figures 3.1 and 3.2 for certain θ values.

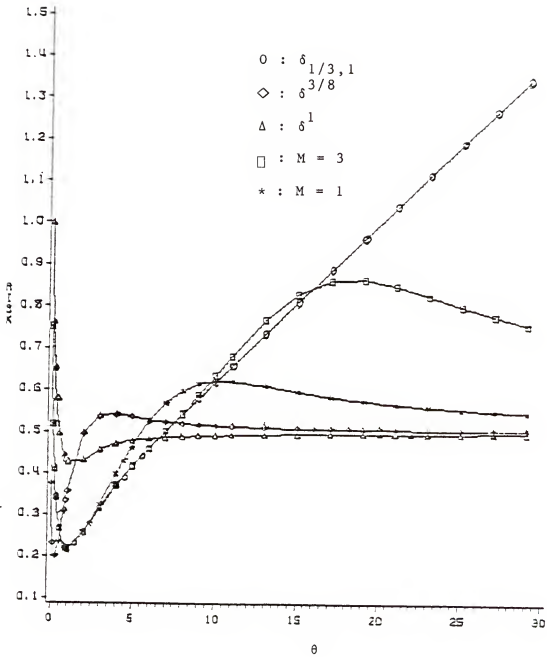


Figure 3.1 Risk of the limited translation
rule $\delta_{M,1/3,1}^{3/8}$

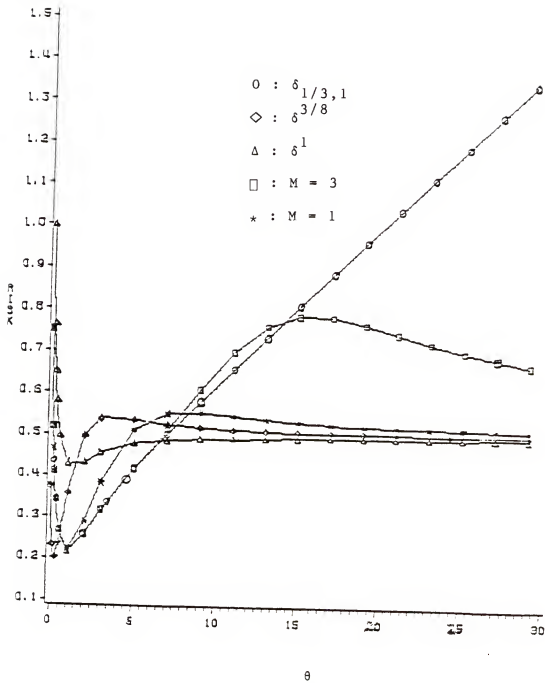


Figure 3.2 Risk of the limited translation
rule $\delta_{M,1/3,1}^1$

Table 3.1 Values for the risk of the limited
translation rule $\delta_{M,1/3,1}^C$

C		3/8		1	
θ	M	1	3	1	3
0		.75	.75	.75	.75
0.1		.5167	.5167	.5167	.5167
0.2		.4090	.4090	.4090	.4090
0.3		.3414	.3414	.3415	.3415
0.8		.2215	.2215	.2236	.2215
1		.2144	.2143	.2187	.2143
2		.2580	.2556	.2862	.2556
3		.3269	.3150	.3818	.3150
4		.3985	.3681	.4597	.3682
5		.4667	.4151	.5107	.4161
8		.5978	.5438	.5533	.5576
10		.6198	.6340	.5485	.6552
15		.5959	.8293	.5319	.7855
21		.5661	.8483	.5218	.7434
25		.5544	.8000	.5180	.7022
31		.5430	.7367	.5143	.6587
51		.5254	.6367	.5085	.5922

3.3 The RSL of $\delta_{M,\alpha,k}^C$

The RSL of $\delta_{M,\alpha,k}^C$ is defined by

$$\text{RSL}(\Pi, \delta_{M,\alpha,k}^C) = \frac{r(\Pi, \delta_{M,\alpha,k}^C) - r(\Pi, \delta_{\alpha,k})}{r(\Pi, \delta^C) - r(\Pi, \delta_{\alpha,k})}, \quad (3.3.1)$$

where Π denotes the $\text{gamma}(\alpha, k)$ prior. The expression (3.3.1) can be interpreted as the proportion of the Bayes risk improvement over δ^C that is sacrificed by the use of $\delta_{M,\alpha,k}^C$ instead of the Bayes estimator with respect to the prior Π .

THEOREM 3.3.1. $r(\Pi, \delta_{M,\alpha,k}^C) \rightarrow r(\Pi, \delta_{\alpha,k})$ as $M \rightarrow \infty$.

Proof. Assume $(M_0 - C + b)/(1 - a) = d$ and $(M_1 - C + b)/(1 - a) = d + 1$ where d is a positive integer. Consider

$$\begin{aligned} & R(\theta, \delta_{M_1,\alpha,k}^C) - R(\theta, \delta_{M_0,\alpha,k}^C) \\ &= \sum_{x=d+1}^{\infty} \left\{ (a-1) + \theta \log\left(\frac{x+C-M_0}{x+C-M_1}\right) \right\} \frac{\exp(-\theta)\theta^x}{x!} \\ &= \sum_{x=d+1}^{\infty} \left\{ (a-1) + \theta \log\left(1 + \frac{1-a}{x+C-M_1}\right) \right\} \frac{\exp(-\theta)\theta^x}{x!}. \end{aligned}$$

Then,

$$r(\Pi, \delta_{M_1,\alpha,k}^C) - r(\Pi, \delta_{M_0,\alpha,k}^C)$$

$$\begin{aligned}
&= \int_0^{\infty} (R(\theta, \delta_{M_1, \alpha, k}^C) - R(\theta, \delta_{M_0, \alpha, k}^C)) \Pi(\theta) d\theta \\
&= \int_0^{\infty} \sum_{x=C+1}^{\infty} \left((a-1) + \theta \log\left(1 + \frac{1-a}{x+C-M_1}\right) \right) \frac{\exp(-(1+\alpha)\theta) \theta^{x+k-1} \alpha^k}{x! \Gamma(k)} d\theta \\
&= \sum_{x=d+1}^{\infty} \frac{\alpha^k}{\Gamma(k)} \left(\int_0^{\infty} (a-1) \exp(-(1+\alpha)\theta) \theta^{x+k-1} d\theta \right) + \\
&\quad \sum_{x=d+1}^{\infty} \frac{\alpha^k}{\Gamma(k)} \log\left(1 + \frac{1-a}{x+C-M_1}\right) \left(\int_0^{\infty} \exp(-(1+\alpha)\theta) \theta^{x+k} d\theta \right) \\
&= \sum_{x=d+1}^{\infty} \left((a-1) + \log\left(1 + \frac{1-a}{x+C-M_1}\right) \frac{x+k}{1+\alpha} \right) \frac{\alpha^k \Gamma(x+k)}{\Gamma(k)(1+\alpha)^{x+k}}.
\end{aligned}$$

Since

$$\begin{aligned}
&\log\left(1 + \frac{1-a}{x+C-M_1}\right) (x+k) (1/(1+\alpha)) \\
&< a(1-a)(x+k)/(x+C-M_1), \quad x \geq d+1, \\
&= (1-a)a(d+1+k+x)/(d+1+C-M_1+x), \quad x = 0, 1, 2, \dots \\
&= (1-a)a(d+1+k+x)/(ad+a+b+x), \quad x = 0, 1, 2, \dots \\
&= (1-a)(ad+a+ak+ax)/(ad+a+b+x), \quad x = 0, 1, 2, \dots \\
&= (1-a)(ad+a+b+ax)/(ad+a+b+x), \quad x = 0, 1, 2, \dots \\
&< 1-a, \text{ we have}
\end{aligned}$$

$$(a-1) + \log\left(1 + \frac{1-a}{x+C-M_1}\right) (x+k)/(1+\alpha) < 0 \text{ when } x \geq d+1.$$

Thus, $r(\Pi, \delta_{M_1, \alpha, k}^C) < r(\Pi, \delta_{M_0, \alpha, k}^C)$ if $M_1 > M_0$. Note that if there exists $M_1 \geq M_0$ such that $[(M_0 - C + b)/(1-a)] = d$ and $[(M_1 - C + b)/(1-a)] = d$, then

$$\begin{aligned}
&r(\Pi, \delta_{M_1, \alpha, k}^C) - r(\Pi, \delta_{M_0, \alpha, k}^C) \\
&= \sum_{x=d+2}^{\infty} \left((a-1) + \log\left(1 + \frac{1-a}{x+C-M_1}\right) \frac{x+k}{1+\alpha} \right) \frac{\alpha^k \Gamma(x+k)}{\Gamma(k)(1+\alpha)^{x+k}} < 0.
\end{aligned}$$

Therefore, $r(\Pi, \delta_{M,\alpha,k}^C) + r(\Pi, \delta_{\alpha,k})$ as $M \rightarrow \infty$, since $\delta_{M,\alpha,k}^C \xrightarrow{a.s.} \delta_{\alpha,k}$ as $M \rightarrow \infty$. The proof is complete.

Next we calculate RSL of $\delta_{M,\alpha,k}^C$. Rewrite (3.3.1) as

$$\begin{aligned} & \text{RSL}(\Pi, \delta_{M,\alpha,k}^C) \\ &= 1 - \frac{r(\Pi, \delta^C) - r(\Pi, \delta_{M,\alpha,k}^C)}{r(\Pi, \delta^C) - r(\Pi, \delta_{\alpha,k})}. \end{aligned} \quad (3.3.2)$$

Now

$$\begin{aligned} & r(\Pi, \delta^C) - r(\Pi, \delta_{M,\alpha,k}^C) \\ &= C - \int_0^\infty \left(\sum_{x=0}^d (ax+b) + \sum_{x=d+1}^\infty (x+C-M) \right) \frac{\exp(-(1+\alpha)\theta) \theta^{x+k-1} \alpha^k}{x! \Gamma(k)} d\theta + \\ & \quad \int_0^\infty \left(\sum_{x=0}^\infty \log\left(\frac{ax+b}{x+C}\right) + \sum_{x=d+1}^\infty \log\left(\frac{x+C-M}{x+C}\right) \right) \frac{\exp(-(1+\alpha)\theta) \theta^{x+k} \alpha^k}{x! \Gamma(k)} d\theta + \frac{k}{\alpha} \\ &= C + k/\alpha - E[f(X)] + E[g(Y)], \end{aligned} \quad (3.3.3)$$

where

$$f(x) = \begin{cases} ax + b, & x \leq d \\ x + C - M, & x > d \end{cases},$$

$$g(y) = \begin{cases} \log((ay+b)/(y+C)), & y \leq d \\ \log((y+C-M)/(y+C)), & y > d. \end{cases},$$

X has a negative binomial distribution $NB(k, \alpha/(1+\alpha))$ of the form

$$P(X = x) = \frac{\Gamma(x + k)}{\Gamma(x + 1)\Gamma(k)} \left(\frac{\alpha}{\alpha + 1}\right)^k \left(\frac{1}{\alpha + 1}\right)^x, \text{ and the}$$

probability function of Y is

$$P(Y = y) = \frac{\Gamma(y + 1 + k)}{\Gamma(y + 1)\Gamma(k)} \left(\frac{\alpha}{\alpha + 1}\right)^k \left(\frac{1}{\alpha + 1}\right)^{y+1}. \text{ Similarly,}$$

$$r(\Pi, \delta^C) - r(\Pi, \delta_{\alpha, k}) = C + E[h(Z)], \quad (3.3.4)$$

where $h(z) = \log((az+b)/(z+C))$ and the probability function of Z is

$$P(Z = z) = \frac{\Gamma(z + 1 + k)}{\Gamma(z + 1)\Gamma(k)} \left(\frac{\alpha}{\alpha + 1}\right)^k \left(\frac{1}{\alpha + 1}\right)^{z+1}.$$

Combineing (3.3.3) and (3.3.4), we get

$$RSL(\Pi, \delta_{M, \alpha, k}^C) = 1 - \frac{C + \frac{k}{\alpha} - E[f(X)] + E[g(Y)]}{C + E[h(Z)]}.$$

Also note that $E[f(X)] < k/\alpha$, so that

$$RSL(\Pi, \delta_{M, \alpha, k}^C) < \frac{E[h(Z)] - E[g(Y)]}{C + E[h(Z)]}.$$

Figures 3.3 and 3.4 are plots of the relative saving loss against values of M of the limited translation rule with the value $k = 1$ and the values $\alpha = 1/3, 1$ and 3 . The two graphs correspond to two values of C , namely, $C = 1$ and $3/8$. The larger value of M , the closer the compromise estimator is to the Bayes estimator, and accordingly the RSL value is smaller. This is clearly depicted in Figures 3.3 and 3.4. Table 3.2 presents several RSL values in the above figures.

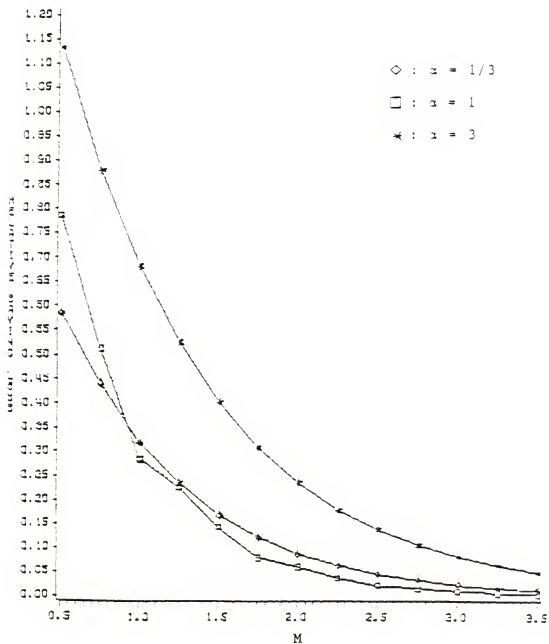


Figure 3.3 Relative saving loss of $\delta_{M,\alpha,1}^{3/8}$

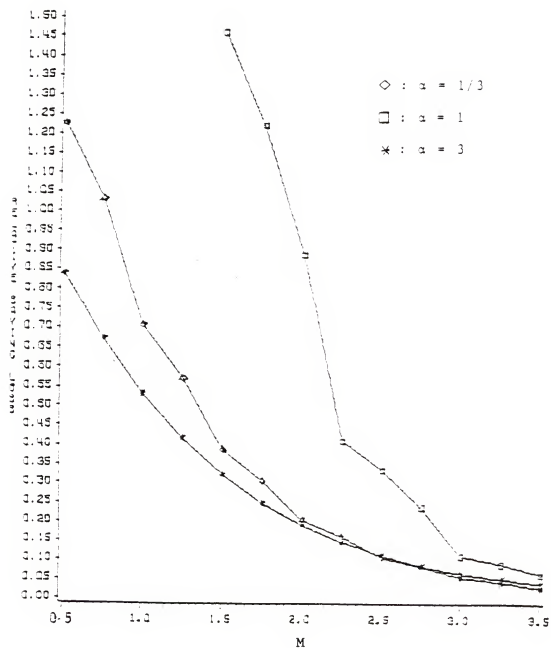


Figure 3.4 Relative saving loss of $\delta_{M,\alpha,1}^1$

Table 3.2 Values for the Relative Saving loss of $\delta_{M,\alpha,l}^C$

C		3/8			1		
M	α	1/3	1	3	1/3	1	3
0.5		1.1318	.7855	.5856	.8388	5.1654	1.2274
0.75		.8793	.5113	.4418	.6728	4.7971	1.0314
1.0		.6792	.2836	.3163	.5306	4.1751	.7066
1.25		.5226	.2244	.2345	.4140	3.1622	.5714
1.50		.4009	.1410	.1667	.3208	1.4564	.3849
1.75		.3070	.0786	.1225	.2473	1.2178	.3052
2.0		.2348	.0608	.0867	.1900	.8855	.2041
2.25		.1795	.0376	.0633	.1457	.4083	.1600
2.5		.1374	.0211	.0447	.1115	.3338	.1066
2.75		.1052	.0161	.0325	.0853	.2375	.0829
3.00		.0808	.0098	.0229	.0653	.1103	.0551
3.25		.0622	.0055	.0166	.0500	.0888	.0426
3.5		.0481	.0042	.0117	.0383	.0624	.0283

CHAPTER FOUR SUMMARY

In Chapter Two, we characterize admissible linear estimators of Poisson means with the form $\mathbf{X}\mathbf{C} + \mathbf{b}$, where $\mathbf{C}$ is a known diagonal matrix and $\mathbf{b}$ is a known vector, under entropy loss. Also, we present some interesting admissible proper Bayes estimators by using hierarchical priors. We point out also in this chapter that Bayes estimators can be viewed as ridge estimators.

In Chapter Three, we construct compromise estimators between generalized Bayes and Bayes estimators using the "Limited Translation Rules" proposed by Efron and Morris (1971). The risk and Bayes risk performance of these compromise estimators will be discussed in this chapter. Also, the "Relative Saving Loss" from Efron and Morris (1973) will be used to compute the proportion of Bayes risk improvement over the rival generalized Bayes estimator that is sacrificed by the use of a compromise estimator instead of a Bayes estimator when the prior is true.

BIBLIOGRAPHY

- Anscombe, F. J. (1948). The transformation of Poisson binomial and negative binomial data. Biometrika, 35, 246-254.
- Anscombe, F. J. (1956). On estimating binomial response relations. Biometrika, 43, 461-464.
- Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility. Ann. Math. Statist., 22, 22-42.
- Brown, L. D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. Ann. Math. Statist., 39, 29-48.
- Brown, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary-value problems. Ann. Math. Statist., 42, 855-903.
- Brown, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. Ann. Statist., 9, 1289-1300.
- Brown, L. D., and Farrell, R. H. (1985). All admissible linear estimators of a multivariate Poisson mean. Ann. Statist., 13, 282-294.
- Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. Statistical Decision Theory and Related Topics III, 1, 205-230, Academic Press, New York.
- Clevenson, M. L., and Zidek, J. V. (1975). Simultaneous estimation of the mean of independent Poisson laws. J. Amer. Statist. Assoc., 70, 698-705.
- Cohen, A. (1966). All admissible linear estimates of the mean vector. Ann. Math. Statist., 37, 458-463.

- Dey, D. K., and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. Ann. Statist. 13, 1581-1591.
- Dey, D. K., and Srinivasan, C. (1986). On estimation of Poisson Logits. To appear in Comm. Statist., Theory and Methods.
- Dey, D. K., Ghosh, M. and Srinivasan, C. (1986). Simultaneous estimation of parameters under entropy loss. To appear in J. Stat. Plan. and Inf.
- Efron, B., and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators- Part 1: the Bayes case. J. Amer. Statist. Assoc., 66, 807-815.
- Efron, B., and Morris, C. (1973). Stein's estimation rule and its competitors- an empirical Bayes approach. J. Amer. Statist. Assoc., 68, 117-130.
- Ghosh, M. (1983). Estimates of multiple Poisson means: Bay and empirical Bayes. Statistics and Decisions, 1, 183-195.
- Ghosh, M., Hwang, J. T. and Tsui, K. (1983). Construction of improved estimators in multiparameter estimation for discrete exponential families (with discussion). Ann. Statist., 11, 351-376
- Ghosh, M., and Parsian, A. (1981). Bayes minimax estimation of multiple Poisson parameters. J. Mult. Anal., 11, 280-288
- Girshick, M. A., and Savage, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. Proc. 2nd Berk. Symp. Math. Statist. Prob., 1, 53-73, Univ. of California Press, Berkeley.
- Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix. J. Mult. Anal., 7, 374-385.
- Haff, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. Ann. Statist., 7, 1264-1276.
- Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. Ann. Statist., 8, 586-597.

- Haff, L. R. (1982). Identities for the Wishart distribution with applications to regression. Sankhya B, 44, 245-258.
- Haldane, J. B. S. (1956). The estimation and significance of the logarithm of a ratio of frequencies. Ann. Hum. Gen., 20, 309-311.
- Hoadley, B. (1981). Quality management plan (QMP). Bell System Technical Journal, 60, 215-274.
- Hodges, J. L., Jr., and Lehmann, E. L. (1951). Some applications of the Cramer-Rao inequality. Proc. 2nd Berk. Symp. Math. Statist. Prob., 1, 13-22, Univ. of California Press, Berkeley.
- Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.
- Hudson, H. M., and Tsui, K. W. (1981). Simultaneous Poisson estimators for a prior hypotheses about means. J. Amer. Statis. Assoc., 76, 182-187.
- Hwang, J. T. (1982). Improving upon inadmissible estimators in discrete exponential families with applications to Poisson and negative binomial cases. Ann. Statist., 10, 857-867.
- Ighodaro, A. O. (1980). Ridge and James-Stein methods for contingency tables. Ph.D. Thesis, School of OR/IE, Cornell University, Ithaca.
- Ighodaro, A. O., and Santner, T. (1982). Ridge type estimators of Multinomial cell probabilities. Statistical Decisions Theory and Related Topics III, 2, 31-53, Academic Press, New York.
- Ighodaro, A., Santner, T. and Brown, L. D. (1982). Admissibility and complete class results for the multinomial estimation problem with entropy and squared error loss. J. Multivariate Anal., 12, 469-479.
- James, W., and Stein, C. (1961). Estimation with quadratic loss. Proc. 4th Berk. Symp. Math. Stat. Prob., 1, 361-367, Univ. of California Press, Berkeley.

- Johnson, B. M. (1971). On the admissible estimators for certain fixed sample binomial problems. Ann. Math. Statist., 42, 1579-1587.
- Karlin, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist., 29, 406-436.
- Kullback, S. (1959). Information theory and statistics. New York: Wiley.
- Matsumura, E. M., and Tsui, K. W. (1982). Stein type Poisson estimators in audit sampling. J. Accounting Research, 20, 162-170.
- Peng, J. C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions. Technical Report #78. Department of Statistics, Stanford University.
- Rolph, J. E., Chaiken, J. M. and Houchens, R. L. (1981). Methods for estimating crime rates of individuals. Rand corporation, Technical Report #R-2730-NIJ.
- Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Prod. Third Berk. Symp. Math. Statist. Prob., 1, 197-206, Univ. of California Press, Berkeley.
- Stein, C. (1973). Estimation of the mean of a multivariate distribution. Proc. Prague Symp. Asymtotic Statist., 345-381.
- Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Statist., 42, No.1, 385-388.
- Sutherland, M., Holland, P. and Feinberg, S. (1974). Combining Bayes and frequency approaches to estimate a multinomial parameter, Studies in Bayesian Econometrics and Statistics, S. Feinberg and A. Zellner (eds.). North Holland, Amsterdam.
- Tsui, K. W. (1986). Further developments on the robustness of Clevenson-Zidek-Type means estimators. J. Amer. Statist. Assoc., 81, 176-180

BIOGRAPHICAL SKETCH

Ming-Chung Yang was born in Chiayi, Taiwan, on August 11, 1952. He received a Bachelor of Science degree in mathematics from Tamkang University, in 1975. After one year, he entered the National Taiwan Normal University and received a Master's in mathematics in 1978. From 1975 to 1978, he worked as a teaching assistant in Tamkang and National Taiwan Normal University. He enrolled in the Department of Statistics at the University of Florida in August, 1983. He is expected to receive the degree of Doctor of Philosophy in August 1986. While at the University of Florida, he has been employed by the Department of Statistics as a research assistant under Professor Alan Agresti of University of Florida on a research grant from the National Institute of Health. He was also employed as a teaching assistant at the Department of Statistics and a statistical consultant at the Institute of Food and Agriculture Sciences Statistics Unit.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Malay Ghosh

Malay Ghosh, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Murali Rao

Murali Rao
Professor of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ronald H. Randles

Ronald Randles
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and it fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andrew Rosalsky

Andrew Rosalsky
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Jeffrey Witmer
Assistant Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted for partial fulfillment of the requirements of the degree of Doctor of Philosophy.

August, 1986

Dean, Graduate School